

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 22, 2004, 21:22:24 ; Search time 5080 Seconds
(without alignments)
4873.179 Million cell updates/sec

Title: US-09-916-849a-1
Perfect score: 4369
Sequence: 1 MGLPRLASLLLLQVWLQ.....NEWGSRFKKLADMYGGEDD 829

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 550265578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_plus_p2n.model -DRV=xlh
-Q=/cgn2_1/USPTO.spool.h/US09916849/runat_21092004_145803_7018/app_query.fasta_1.967
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09916849_CGN_1_1_3609@runat_21092004_145803_7018 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hctc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hctc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_ges_hum:*
18: em_ges_inv:*
19: em_ges_pln:*
20: em_ges_vrt:*
21: em_ges_fun:*
22: em_ges_mam:*
23: em_ges_mus:*
24: em_ges_pro:*
25: em_ges_rod:*
26: em_ges_pbg:*
27: em_ges_vrl:*
28: gb_gesl:*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4310	98.6	3632	11	BC014462	BC014462 Homo sapi
2	3564.5	81.6	3971	11	AK031265	AK031265 Mus muscu
3	3560	81.5	3233	11	AK045041	AK045041 Mus muscu
4	2351	53.8	4415	11	AK076369	AK076369 Mus muscu
5	1701	38.9	3136	11	AK032826	AK032826 Mus muscu
6	1486	34.0	1016	12	BM923882	BM923882 AGENCOURT
7	1459	33.0	873	14	CA453911	CA453911 AGENCOURT
8	1440	33.0	887	14	CA455046	CA455046 AGENCOURT
9	1417.5	32.4	1049	13	BQ936686	BQ936686 AGENCOURT
10	1408	32.2	912	13	BQ927154	BQ927154 AGENCOURT
11	1382.5	31.6	846	13	BQ690885	BQ690885 AGENCOURT
12	1382.5	31.6	918	14	CA454560	CA454560 AGENCOURT
13	1379.5	31.6	912	13	BQ928283	BQ928283 AGENCOURT
14	1350	30.9	875	14	CA489015	CA489015 AGENCOURT
15	1343	30.7	913	13	BQ676796	BQ676796 AGENCOURT
16	1342	30.7	836	12	BM045728	BM045728 603624563
17	1329	30.4	865	13	BU543746	BU543746 AGENCOURT
18	1324.5	30.3	999	12	BM804940	BM804940 AGENCOURT
19	1321	30.2	892	14	CA489338	CA489338 AGENCOURT
20	1312	30.0	894	14	CA489124	CA489124 AGENCOURT
21	1310.5	30.0	899	14	CA454360	CA454360 AGENCOURT
22	1302	29.8	870	10	BE378969	BE378969 601237596
23	1293.5	29.6	773	10	BE740340	BE740340 601594641
24	1277	29.2	1122	12	BM545620	BM545620 AGENCOURT
25	1268	29.0	958	13	BQ886447	BQ886447 AGENCOURT
26	1264.5	28.9	927	14	CA489357	CA489357 AGENCOURT
27	1261	28.9	1001	13	BQ072354	BQ072354 AGENCOURT
28	1250.5	28.6	888	14	CA487856	CA487856 AGENCOURT
29	1238	28.3	745	14	CB240552	CB240552 UT-CF-FN0
30	1231	28.2	727	12	BM047867	BM047867 603618920
31	1219.5	27.9	769	12	BG403092	BG403092 602418911
32	1219.5	27.9	778	12	BM049731	BM049731 603624246
33	1212	27.7	901	14	CA454213	CA454213 AGENCOURT
34	1212	27.7	1006	13	BU168948	BU168948 AGENCOURT
35	1210.5	27.7	916	13	BU179183	BU179183 AGENCOURT
36	1206	27.6	703	12	BM715802	BM715802 UT-E-EJ0-
37	1199	27.4	1040	12	BM909817	BM909817 AGENCOURT
38	1192	27.3	880	14	CB992509	CB992509 AGENCOURT
39	1172.5	26.8	835	14	CB996355	CB996355 AGENCOURT
40	1151	26.3	761	10	BE741190	BE741190 601593814
41	1140.5	26.1	781	14	CK032635	CK032635 AGENCOURT
42	1108	25.4	949	13	BQ679513	BQ679513 AGENCOURT
43	1104.5	25.3	757	14	CK128630	CK128630 AGENCOURT
44	1103.5	25.3	734	14	CD372275	CD372275 UT-R-G00-
45	1086.5	24.9	906	13	BU145465	BU145465 AGENCOURT

ALIGNMENTS

RESULT 1
BC014462
LOCUS BC014462 3632 bp mRNA linear HTC 19-NOV-2003
DEFINITION Homo sapiens cadherin 3, type 1, P-cadherin (placental), mRNA (cdna
clone IMAGE:4870356), containing frame-shift errors.
ACCESSION BC014462
VERSION BC014462.1 GI:15680220
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3632)

Db 1408 GCAGTGGCCATGAGGTGACAGGCTGACGGTCACTGATCTGACAGCGCCCAACTCACCA 1467
Qy 361 AlaTrpAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db 1468 GCGTGGCGTCCCACTACCTTATCATGGCGGTGACGACGGGACCATTTTACCATCAC 1527
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgGlyGlyLeuAspPheGluAla 400
Db 1528 ACCACCTGAGAGCAACCAAGGATCTCTGACAAACCAAGGAGGTTTGGATTGTGAGGCC 1587
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1588 AAAAACCCAGCACACCTGTAGTTGAAGTGACCAACGAGGCCCTTTTGTGCTGAGACTC 1647
Qy 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
Db 1648 CCAACCTCCACAGCCACCTAGTGGTCCACGTCGAGGATGTGAATGAGGACCTGTGTGTT 1707
Qy 441 ValProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1708 GTCCCACTCCAAAGTGGTTGAGGTCAGAGGGATCCCACTGGGAGGCTGTGTGT 1767
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1768 GTCTACACTGCAGAGACCCCTGCAAGAGGAAATCAAAGATCAGCTACCGCATCCTGAGA 1827
Qy 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1828 GACCCAGAGGTGGTTCAGCATGGACCCAGACAGTGGGAGGTACAGCTGTGGGCACC 1887
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1888 CTCGACCGTGAGGATGAGCATGTTGTGAGGAAACAACATCTATGAAGTCATGTGTC 1947
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeuAsp 540
Db 1948 ATGGAACAATGGAGCCCTCCACCATCGCAGCGGAAACCTTCTGCTAACACTGATGAT 2007
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 2008 GTCAACGACCATGGCCCATGCTCCCTGAGCCCGCTGAGTACCATCAGTCAACCAAGCCCT 2067
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 2068 GTGCGCCAGGTGCTGAACATCAAGCAAGGACCTGTCTCCCAACACCTCCCTTTCCAG 2127
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 2128 GCCAGCTCACAGTACATCAGACATCTACTGACGCGAGAGGTCAACGAGGAGGTGAC 2187
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 2188 ACAGTGGTCTTGTCCCTGAAGAAGTTCCTGAAGCAGGATACATATGACGTGCACCTTCT 2247
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 2248 CTGTCTGACCATGGCAACAAAGACGAGCTGACGGGTGATCAGGGCCACGTGTGCGACATGC 2307
Qy 641 HisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValLeu 660
Db 2308 CATGGCCATGTCGAACCTGCCCTGGACCTCGAAGGAGGTTTCTATCTCCTCTGCTG 2367
Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2368 GGGGTGCTCCTGGCTCTGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2427
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2428 CGGAGATCAGAGAGCCCT 2487
Qy 701 TyrGlyGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2488 TATGGCGAAGAGGGGGTGGCGAAGAGGACCAAGGACTATGACATCACCCAGCTCCACCGA 2547

Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db 2548 GGTCTGGAGGCCAGGCCGAGGTGGTCTCTCGCAATGACGTGGCACCAACCATCATCCCG 2607
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db 2608 ACACCCATGTACCGTCTCTAGGCCAGCAACCCAGATGAATCGGCACTTTTATATTGA- 2666
Qy 760 ----- 760
Db 2667 GGGGAGGGAGAGAGGGGCTCACAGAGAGAAATGGAGGCTTGCGCTGGCAAGAGGAG 2726
Qy 761 -----AsnLeuLysAlaAlaAsnThr 767
Db 2727 GACCGCGCGCTCCTAACTACCTGTTCTGTGTGCGCAGAAACCTGAAGCGCGCTAACACA 2786
Qy 768 AspProThrAlaProProTyrAspThrLeuLeuValPheAspTyrGlyGlySerGlySer 787
Db 2787 GACCCACAGCCCGCCCTTACGACACCTCTTGTGTGTGACTATGAGGCGCGGTCTCC 2846
Qy 788 AspAlaAlaSerLeuSerSerLeuThrSerSerAlaSerAspGlnAspGlnAspTyrAsp 807
Db 2847 GACGCGCGTCCCTGAGTCCCTCACCTCTCCGCTCCGACCAAGCAAGATTACGAT 2906
Qy 808 TyrLeuAsnGluTrpGlySerArgPheLysLysLeuAlaAspMetTyrGlyGlyGlu 827
Db 2907 TATCTGAACGAGTGGGGCAGCCGCTTCAAGAAAGTGGCAGACATGTACGTTGGCGGGAG 2966
Qy 828 AspAsp 829
Db 2967 GACGAC 2972

RESULT 2
AK031265
LOCUS
DEFINITION Mus musculus 13 days embryo forelimb cDNA, RIKEN full-length enriched library, clone:5930439E13 product:cadherin 3, full insert sequence.
ACCESSION AK031265
VERSION AK031265.1 GI:26327166
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
PUBMED 10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
PUBMED 11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, K., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
PUBMED 11076861


```

QY 299 GlnAlaThrAspMetAspGlyAspGlySerThrThrAlaValAlaValGluIle 318
Db 918 CAGCCACACAGATGATGAGAGGGCTCTACACACGACGAGAGCCGTGTGCAATC 977
QY 319 LeuAspAlaAsnAspAsnAlaProMetPheAspProGlnIleTyrGluAlaHisValPro 338
Db 978 CTTGATGCCAACGATACCGCTCCGGAGTTTGACCCGACAGAGTATGAGGCTTGGGTGCCT 1037
QY 339 GluAsnAlaValGlyHisGluValGlnAlaArgLeuThrValThrAspLeuAspAlaProAsn 358
Db 1038 GAGAACGAGATGGGCCATGAGGTACAGAGGCTGACAGTCACTGATCTCGATGTCCCAAC 1097
QY 359 SerProAlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThr 378
Db 1098 TCCCGAGCGTGGGTGCCCTACCATCTGCTGGAGGTGATGATGGGACCATTTCCAC 1157
QY 379 IleThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPhe 398
Db 1158 ATCACCACCTCACCAGAGACCAACCAAGGCTCTGACACCAACCAAGAGGTTTGGATTTT 1217
QY 399 GluAlaLysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeu 418
Db 1218 GAGGCTCAGGACCAACACACCCCTGTATGTAGAGTACCACAGGCTCCCTTTGCAGTG 1277
QY 419 LysLeuProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaPro 438
Db 1278 AAGCTCCGACAGCCACTGCCACCGTGGTGTCTCATGTGAAGATGTCAACGAAGCCCT 1337
QY 439 ValPheValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluPro 458
Db 1338 GTGTTTGTCCACCTTCCAAAGGTCATGTAGGCGCCAGGAAGGCTCTCTATTGGGGAACGT 1397
QY 459 ValCysValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIle 478
Db 1398 GTCTGCATCTATACCGACAGGACCCAGACAGGAGGACCAAGATCAGCTACACCATC 1457
QY 479 LeuArgAspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaVal 498
Db 1458 TCAGAGATGCCAGCAACTGGCTGTGTGGACCCAGACAGTGGTCAGATACTGCCGCA 1517
QY 499 GlyThrLeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetVal 518
Db 1518 GGCATCTTGATCTGTGAGACGACGAGCTTTGGAAACAACATGTCTACGAAGTCATGGTT 1577
QY 519 LeuAlaMetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeu 538
Db 1578 TTGGCCACAGACAGTGGAAACCCCTCCACCAACCGGCACTGGGACCTCTGCTTACACTT 1637
QY 539 IleAspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGln 558
Db 1638 ACTGATCAACAGCACCATGGCCGATCCCTGTGAACCCAGGACAGATCATCTGTAAACCA 1697
QY 559 SerProValArgHisValLeuAsnIleThrAspLysAspLysSerProHisThrSerPro 578
Db 1698 AGCCCTGTGCCTCAAGTGTGAACATCACTGACAGGACCTGTCCCGCAACTCTCCCT 1757
QY 579 PheGlnAlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGlu 598
Db 1758 TTCCAGCCCGCAGCTAACACATGACTCAGATATCTACTGGATGGCAGAGTCAAGAGAA 1817
QY 599 GlyAspThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHis 618
Db 1818 GGNAGACCGTGGCTTGTCCCTGAGAGAGTTCCTGAAACAAGACACGATGATGACTTGAT 1877
QY 619 LeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCys 638
Db 1878 CTTTCTTTCTGACCATGGCAACAGGAAAGCAAGTAAACCATGATCAGGCCACTGTGTGT 1937
QY 639 AspCysHisGlyHisVal---GluThrCysProGlyProTyrLysGlyGlyPheLeuLeu 657
Db 1938 GACTGCCATGGCCAAAGTGTTCATGATGATGATGATGATGATGATGATGATGATGATGAT 1997
QY 658 ProValLeuGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuValLeuVal 677

```

```

Db 1998 CCCATCTCTGGTGTCTCTGGCACTGCTGACCCCTTACTGGCACTCCTCTGTGTGTG 2057
QY 678 ArgLysLysArgLysIleLysGluProLeuLeuLeuProGluLysAspThrArgAspAsn 697
Db 2058 AGGAAGAAGAGAGAGGTCAAGAGCCCTTCTGCTCCAGAGATGACACCGGAGACAA 2117
QY 698 ValPheTyrTyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGln 717
Db 2118 GTCTTCTATTATGGAGAGAGGGTGGTGGTGAAGAGGACAGGACTATGACATCACCAC 2177
QY 718 LeuHisArgGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThr 737
Db 2178 CTCACCGGGGACTGGAGGCCAGGCTGAGGTGTCTCCGAAACGATGTAGTCCCAAC 2237
QY 738 IleIleProThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPhe 757
Db 2238 TTCATCCCAACCCCATGTACCGACCCCGGCCCAACCCAGATGAATCGGAACCTTC 2297
QY 758 IleIleGluAsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeu 777
Db 2298 ATCATCGAAGACCTGAAGCTGCCAACAACACTGACCTACTGCCCGCCCTACGACTCCTG 2357
QY 778 LeuValPheAspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSer 797
Db 2358 CTGGTTTTTGACTACGAGGGCAGCGGCTCTGATCGCGCTCCCTGAGCTCCTCACCAC 2417
QY 798 SerAlaSerAspGlnAspGlnAspTyrAspTyrIleAsnGluTrpGlySerArgPheLys 817
Db 2418 TCCGCTCCCGACAGGATCAGGACTACAACTACACTTAACGAGTGGGAAGTCGATTCA 2477
QY 818 LysLeuAlaAspMetTyrGlyGlyGlyLysAspAsp 829
Db 2478 AAACCTGGCGGACATGTATGTTGGCGGTGAGGATGAC 2513

RESULT 3
AK045041 3233 bp mRNA linear HTC 20-SEP-2003
LOCUS Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
DEFINITION enriched library, clone:BI30021C20 product:cadherin 3, full insert
sequence.
ACCESSION AK045041
VERSION AK045041.1 GI:26337020
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
2 Carninci, P. and Hayashizaki, Y.
Normalizing and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuda, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

```


Db 818 AGCCAAGAGCGAAGAGCGACACGACCTCATGTTCACCATCCATAAAGACGCGGAACC 877
 Qy IleserValIleserSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIle 298
 Db 878 ATTAGCGTCATATCCAGTGGCTGGACCGAGAGAAAGTCCCTGAGTACACACTGACCGTC 937
 Qy GlnIleThrAspMetAspGlyAspGlySerThrThrAlaValAlaValGluIle 318
 Db 938 CAGGCCACAGACATGGATGGAGAGGGCTCTACCAACGACGCGAGAGGCGGTGTGTGCAATC 997
 Qy LeuAspAlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValPro 338
 Db 998 CTTGATGCCAAGCATAAGCTCCGAGTTTGGCCGCGAGAGTATGAGGCTTGGTGCCCT 1057
 Qy GluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsn 358
 Db 1058 GAGAACCAAGTGGGCCCATGAGGTACAGAGGCTGACAGGCTGATCTCGATGTCCCAAC 1117
 Qy SerProAlaIleArgAlaThrTyrLeuIleMetGlyValAspAspGlyAspHisPheThr 378
 Db 1118 TCGCCAGCGTGGGTGCCACTTACCACATCTGTGGAGGTGATGATGGGACCATTTCCACC 1177
 Qy IleThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPhe 398
 Db 1178 ATCACCACCTCACCAGACGACCAACCAAGGCGCTCTGACACCAAGAGGCTTTGGATTIT 1237
 Qy GluAlaLysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeu 418
 Db 1238 GAGCTCAGGACCAACACACCCCTGTATGTAGAAGTACCAACAGAGCTCCCTTTGCGATG 1297
 Qy LysLeuProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaPro 438
 Db 1298 AAGCTCCGACAGCCACTGCCCGTGGTGTCCATGTGAAGATGTCAACGAGGCCCT 1357
 Qy ValPheValProSerLysValValGluValGlnGluGlyIleProThrGlyGluPro 458
 Db 1358 GTCTTTTGTTCACCTTCCAGAGTCAITGAGGCCGAGGAAGGCATCTCTATTGGGAATG 1417
 Qy ValCysValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIle 478
 Db 1418 GTCTGATCTATACCGCACAGGACCCAGACAGAGGAGCAGAGATCAGCTACACCATC 1477
 Qy LeuArgAspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaVal 498
 Db 1478 TCAGAGATCCAGCCAACTGGCTGTGTGGACCCAGACAGTGGTCAGATACTGCCGCA 1537
 Qy GlyThrLeuAspArgGluAspGluGlnPheValArgAsnIleTyrGluValMetVal 518
 Db 1538 GGCATCTTGATCGTGAGGACGAGCAGTTGTGAAAAACAATGTCTACGAACTCATGGTT 1597
 Qy LeuAlaMetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeu 538
 Db 1598 TTGGCCACAGACAGTGGAAACCTCCACACCGGCACTGGGACCTCTCTGTACACTT 1657
 Qy IleAspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGln 558
 Db 1658 ACTGATCAACAGCACCATGGCCGATCCCTGTAACCCAGGACAGATCATCATCTGTAAACAA 1717
 Qy SerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerPro 578
 Db 1718 AGCCCTGTGCTCAAGTGTGACATCATCTACAGGACCTGTCCCCCACTCTCCCT 1777
 Qy PheGlnAlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlu 598
 Db 1778 TTCAGGCCCCAGCTAACACATGACTCAGATATCTACTGGATGGCAGAGTACAGAGAAA 1837
 Qy GlyAspThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHis 618
 Db 1838 GGAGACACCGTGGCTTGTCCCTGGAAGAAGTTCCCTGAAACAGACACGATGATGATGAT 1897
 Qy LeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCys 638
 Db 1898 CTTTCTTTCTGTGACCATGGCAACAGGGAACAGCTAACCATGATCAGGGCCACTGTGT 1957

Qy 639 AspCysHisGlyHisVal---GluThrCysProGlyProTyrIleGlyClyPheIleLeu 657
 Db 1958 GACTGCCATGGCAAGTGTCAATGACTGCCCGACCCCTGGAAGGGTGGTTTCATCCTC 2017
 Qy 658 ProValLeuGlyAlaValLeuAlaLeuPheLeuLeuValLeuLeuValLeuVal 677
 Db 2018 CCCATCTGGGTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2077
 Qy 678 ArgLysIleArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsn 697
 Db 2078 AGAAGAGAGAGAGAGTCAAGAGCCCTTCTGCTCCCGAAGAGATGACACGCGAGACAA 2137
 Qy 698 ValPheTyrTyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGln 717
 Db 2138 GTCTTCTATTATGAGAGAGAGGTGTGTGTGAGAGGACCCAGGACTATGACATCACCAA 2197
 Qy 718 LeuHisArgGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThr 737
 Db 2198 CTTCCACCGGGGACTCGAGGCCCTGAGGTGTGTCTCCGAAACGATGTAGTGCAACC 2257
 Qy 738 IleIleProThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPhe 757
 Db 2258 TTCATCCCAACCCCATGTATCCGACCCCGCCGACCAACCCAGATGAATCGGGAACCTC 2317
 Qy 758 IleIleGluAsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeu 777
 Db 2318 ATCATCGAAGAACTGAAGGCTGCCAAACACTGACCTACTGCCCCGCCCTAGGACTCCCTG 2377
 Qy 778 LeuValPheAspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSer 797
 Db 2378 CTGGTTTTTGACTACGAGGCGAGCGGCTCTGATGCCGCTCCCTGAGTCCCTCACCAC 2437
 Qy 798 SerAlaSerAspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLys 817
 Db 2438 TCGGCTCCGACCAAGATCAGGACTACACTTACGAGTGGGAGTGGAGTTCGATTCAAG 2497
 Qy 818 LysLeuAlaAspMetTyrGlyGlyGlyGluAspAsp 829
 Db 2498 AAATGGCGGACATGTATGTGGCGGTGAGGATGAC 2533
 RESULT 4
 AK076369
 LOCUS 4415 bp mRNA linear HTC 18-SEP-2003
 DEFINITION Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched library, clone:473469F18 product:cadherin 1, full insert sequence.
 AK076369
 ACCESSION AK076369.1 GI:26096766
 VERSION AK076369.1
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Carninci,P. and Hayashizaki,Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE 2
 AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE 3
 AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kiteunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 PUBMED
 11076861

THE RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5

The FANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 3136)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saito, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
 Direct Submission
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/
 Location/Qualifiers
 1. 3136

FEATURES

source
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM_DB:6720458H05"
 /db_xref="MG:2396474"
 /db_xref="taxon:10090"
 /clone="6720458H05"
 /sex="male"
 /tissue_type="olfactory duct includes surrounding region"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="12 days embryo"
 96..2816
 /note="unnamed protein product; cadherin 2 (MGD|GI:88355,
 GB|NM_007664, evidence: BLASTN, 99%, match=3135)
 putative"
 /codon_start=1
 /protein_id="BAC28044.1"
 /db_xref="GI:26328611"

CDS

/translation="MKRIAGAPRTLPLLAALQASVTSGEIALCKTGFPEDVVSVA
 LPKDHQEGQLLVKFSNCNRKKQVSESEPADFDVDEGTIVYVRSPLFAEAKF
 LTVQDKETQKQVAVNRSRLETPMEKFEIEIVFPRLKSHGALQORQD
 WVIPLNPENSGQVPOELVIRSRDKNLSRYSVTGADQPTGFIINPISGQ
 LSVTKPLRELIAFLHRAVDINGOVENPIDIVINDMNRPEFLHGVNGSV
 PEGSKPGTYWYVTAIDADPNALNGMLRILLSQAPSPSPNMFINNETGDIITVA

TGLDREKVOQVTLIIQATDMEGNPTGLSNTATAVITVDNDNPPFEFTAMTFYGEVP
 ENRDVTVANLTVTDKQPTFAWNAAYRISGGPTGRFAITDPNSNDGLVTVKPI
 DFETNRMTVAENOVPLAKGIQHPPOSTATVSVTVIDNENPYFAPNPKIROBE
 GLHATWLTTLTAQDDPRYMQNIRVTKLSDPANLWIKIDPVNGOITITVLDRESNV
 KNNYNTATFLASDNGIPMGSGTQIQLIILLDINDNAPQVLPQEAETCETPEPNSINIT
 ALDYIDPNAGPFAFDLPSPVTKIKRNTIRNLNGDPAQNLKIKLEAGIYEVPIII
 TDSGNPPKNSISLRVKVQCDSNDGCTDVRIVAGLGTGAIILLLIILLIIVL
 MFWVMKRRDKERQAKLLIDPEDVDNRNLIKDEGGEGEDQDYDLSQIQQPTDTPV
 DAIKPVGIRLDRPIHAEPQYVPSAAPHGIDIGFINELKAANDPPTAPPVDSLL
 VFDEYEGSGTAGSLSSLSNSSSGGDQDYDYLNDWGPFRFKLADWYGGDD"

ORIGIN

Alignment Scores:	7.39e-169	Length:	3136
Pred. No.:	Score:	Matches:	392
Percent Similarity:	1701.00	Conservative:	127
Best Local Similarity:	42.41%	Mismatches:	289
Query Match:	42.61%	Indels:	112
Db:	11	Gaps:	29
US-09-916-849A-1 (1-829) x AK032826 (1-3136)			
QY 2 GlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGlnCys 21			
DB 111 GGAGCGCGCGG-----ACCTGTCGCGCTTCTGGCGGCTTCTTCAGCG 158			
QY 22 AlaAlaSerGluPro-----CysArgAlaValPheArgGluAlaGluVal 36			
DB 159 TCTGTGAGAGCTTCTGTGAAATTCATTATGCAAGTCTGATTTCTCTGAA--GATGTT 215			
QY 37 ThrLeuGluAlaGlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysVal---Phe 55			
DB 216 TAGAGCGAGTCTTACCGAGGATGTCACGAGGACGAGCGCCCTTCTCATGTGAAATTC 275			
QY 56 MetGlyCys-----ProGlyGlnGluProAlaLeuPheSer 67			
DB 276 AGCAACTGCAATAGAAAAGAAAGTTTCAGTATGAAAGCAGCGAGCAGCAGATTTCAAG 335			
QY 68 ThrAspAsnAspAsp-----PheThrValArgAsn----- 77			
DB 336 GTGAGCAGGAGCAGCGCGGTATGCTGTGAGAAAGTCTCCTCTCATGTGAGAGCAGGCA 395			
QY 78 -----GlyGluThrValGlnGluArg----- 84			
DB 396 AAGTCTCTGATATATGCCAAGACAAAGAAACCCAGGAAAGTGGCAGGTAGCTGTAAAC 455			
QY 85 -----ArgSerLeuLysGluArgAsnProLeuLys--- 94			
DB 456 CTGAGCGCGGAGCCCAACCTGACTGAGGAGCCTATGAGGAACCAATGAAATTTGAAGAA 515			
QY 95 --IlePheProSerLys-----ArgIleLeuArgArgHisLysArgAsp 108			
DB 516 ATAGTATTCCTAGACAACTTCCCAAGCAGCAGTGAGGACTCTACAAAGCAGAGAGAGAC 575			
QY 109 TrpValValAlaProIleSerValProGluAsnGlyLysGlyProPheProGlnArgLeu 128			
DB 576 TGGTGTATCCCGCCCAATCACTTCCAGAAACCTCCAGAGACCCCTTCTCTCAAGAGCTT 635			
QY 129 AsnGlnLeuLysSerAsnLysAspArgThrLysIlePheTyrSerIleThrGlyPro 148			
DB 636 GTCAGATCAGGCTGTATAGAGATAAAACCTTTCCTCGATACACGCTCACTGGGCCA 695			
QY 149 GlyAlaAspSerProGluGlyValPheAlaValGluLysGluThrGlyTrpLeuLeu 168			
DB 696 GGAGCTGACCAAGCCTCCCAACCGGCATCTTCAATTATCAACCCCATCTCAGGACAGCTGTCA 755			
QY 169 LeuAsnLysProLeuAspArgGluIleAlaLysTyrGluLeuPheGlyHisAlaVal 188			
DB 756 GTCAGAAAGCCTCTGGATCGAGAGCTGATAGCCCGGTTTCACTTGAGAGCAGATGCAATG 815			
QY 189 SerGluAsnGlyAlaSerValGluAspProMetAsnIleSerIleLeuThrAspGln 208			
DB 816 GACATCAATGCAATCAAGTGGAGAACCCCATTTGACATGTGATCATGTTATTGACATG 875			

Qy	209	AsnAspHisLysProLysPheThrGlnAspThrPheArgGlySerValLeuGluGlyVal	228
Db	876	AATGATACACAGCCTGAGTTTCTGCACACAGGTTTGGAAATGGGTCTGTTCCAGAGGGATCA	935
Qy	229	LeuProGlyThrSerValMetGlnValThrAlaThrAspGluAspAspAlaIleTyThr	248
Db	936	AAGCCTGGAGCATGTGTGATCACCGTCTACTGCCATTGATCGGATGATCCA--AATGCC	992
Qy	249	TyrAsnGlyValValAlaTySerIleHisSerGlnGluProLysAspProHisAspLeu	268
Db	993	CTGAATGGAATGCTGCGGTACAGATCTGTCCAGGCGGCCACAGACACTTACCCCAAC	1052
Qy	269	MetPheThrIleHisArgSerThrGlyThrIleSerValIleSerSerGlyLeuAspArg	288
Db	1053	ATGTTTACATCAAAATGACACTGGGACATCATCTGTGGCACTTGTCTGGATCGA	1112
Qy	289	GluLysValProGluTyThrLeuThrIleGlnAlaThrAspMetAspGlyAsp----	306
Db	1113	GAGAAAGTGCACAACTAGTATAGTTAATAATTCAAGCCACAGACATGGAAGCAATCCCAC	1172
Qy	307	--GlySerThrThrAlaValAlaValValGluIleLeuAspAlaAsnAspAsnAla	325
Db	1173	TATGGCCTTTCAACACAGCCACAGCCGTATCATCGGTGACAGATGCAATGACAAATCCT	1232
Qy	326	ProMetPheAspProGlnLysTyThrGluAlaHisValProGluAsnAlaValGlyHisGlu	345
Db	1233	CCAGAGTTTACTGCCATGACTTTCTACGGAGAAGTCCTGAGAACAGGGTGGACGTCA	1292
Qy	346	ValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerProAlaTTrpArgAlaThr	365
Db	1293	GTACCCAACTAATCTGTACAGCAAAAGATCAGCCACAGCCGCGCTCGAAATCGGCA	1352
Qy	366	TyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThrThrHisProGluSer	385
Db	1353	TACAGAACTAGTGTGTGAGACCTTACAGGAAGGTTTCCCATCTCTGACAGACCCACAC	1412
Qy	386	AsnGlnGlyIleLeuThrThrArgGlyLeuAspPheGluAlaTyAsnGlnHisThr	405
Db	1413	AATGATGGGTAGTCACAGTGGTAAAAACCAATTGACTTTGAAACGAATAGGATGTTG	472
Qy	406	LeuTyValGluValThrAsnGluAlaProPheValLeuLysLeu-----ProThr	422
Db	1473	CTTACTGTGTGCAGAAAACCAAGTGCCTATTAGCTAAAGCATTCAGCACCCCTCAG	1532
Qy	423	SerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPheValPro	442
Db	1533	TCGACAGCCACTGTGTCTGTGACAGTTATTGATGTCATGAAATCTTATTTGCCCCA	1592
Qy	443	ProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCysValTy	462
Db	1593	AATCCTAAAAATCATTCGCCAAGGAAGAGCGCTCCACGACGATCACCATGTCGACCA	1652
Qy	463	ThrAlaGluAspProAspLys--GluAsnGlnLysIleSerTyArgIleLeuArgAsp	481
Db	1653	ACTGCTCAGACCCCGATCGATATATGCCACAGNATATCAGATACACAAATTGTCTG	1712
Qy	482	ProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThrLeu	501
Db	1713	CCTGCCAACTGGCTGAAAATAGACCCCGTGAATGGGCAGATCACTACTATTGCGGTT	1772
Qy	502	AspArgGluAspGluGlnPheValArgAsnAsnIleTyThrGluValMetValLeuAlaMet	521
Db	1773	GACAGAAATCCCAAT--GTAAAAACAACTATTAATGCTACCTTCCCTTGCTTCT	1829
Qy	522	AspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuThrLeuIleAspVal	541
Db	1830	GACAATGGAATCCGCGCTATGAGTGGGACAGGAACACTGCAAAATCTATTACTTGAT	1889
Qy	542	AsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerProVal	561
Db	1890	AATGACAAACCCCTCAGGTGTTTACCTCAAGAGCGGAGACCTGT--GAAATCCAGAA	1946

Qy	562	ArgHisValLeuAenlleThr-----AspLysAspLeuSerProHisThrSerProPhe	579
Db	1947	CCCAACCACTAATTAACACACACAGCACTTGAATATACATACACCAACCGCCGCCGCTTC	2006
Qy	580	GlnAlaGlnLeu-----ThrAspAspSerAspIleTyrTyrThrAlaGlu	594
Db	2007	CGGTTTGATCTTCCTCTATCTCCAGTGCNCTATTAAAGAAAC-----TGGACC-----	2054
Qy	595	ValAsnGlu-----GluGlyAspThrValValLeuSerLeuLys-----LysPheLeuLysGln	612
Db	2055	ATCAACCGCGCTTAATGTGTGATTTTGGCTCAGCTCAATTAAGATAAAATTTTGGGAAGCT	2114
Qy	613	AspThrTyrAspValHisLeuSerLeuSerAspHisSgLYAsn-----LysGluGlnLeu	630
Db	2115	GGTATCTATGAAGTTCATCCATCATATCACAGATTACAGGAATCCCCCAAGTCCAACTT	2174
Qy	631	ThrValIleArgAlaThrValCysAspCysHisSgLYHis-----ValIleu	645
Db	2175	TCCATCTCGCTGTGAAAGTTTGTCAAGTGTGACTCCAATGCAGACTGCACGGACGCTGAC	2234
Qy	646	ThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeuGlyAlaValLeuAla	665
Db	2235	AGGATCGTGGGTGCAGGCTTTGGC-----ACGGGGCGCATCATCGCT	2276
Qy	666	LeuLeuPheLeuLeuValLeuLeuLeuVal-----	677
Db	2277	ATCCTTCTGTGATCATCATCTCTCTGATCCTTGTTCTCATTTGTTGGTATGGATGAAA	2336
Qy	678	-----ArgLysLysArgLysIleLysGluProLeuLeuLeuProGluLysAspThrArg	695
Db	2337	CGGCGGGATAAAGACGCCCAAGCCACGACGCTTTTAATGACCCAGAAAGATGATGAAGA	2396
Qy	696	AspAsnValPheTyrTyrGlyGluGluGlyGlyGluGluAspGlnAspTyrAspIle	715
Db	2397	GATATATATTGAATATGATGAAGAAGTGGAGAGAGAGAACCCAGGACTATGACTTG	2456
Qy	716	ThrGlnLeuHisArgGlyLeuGluAlaAArgProGluVal-----ValLeu	730
Db	2457	AGCCAGCTCCAGCAACACAGATACTGTGGAGCTCATGCCATCAAGCCCGTGGGAATCAGA	2516
Qy	731	ArgAsnAspValAlaProThrIlelleProThrProMetTyrArgProArgProAla-	749
Db	2517	CGGTACAGCAGAGGCTCT---ATCCATGCTGAGCCACAGTACCACGTCCTCGATCCGAC	2573
Qy	750	---AsnProAspGluIleGlyAsnPheIlelleGluAsnLeuLysAlaAlaAsnThrAsp	768
Db	2574	CCACACCTCGGGATATTGGGACTTTCATTAATGAGGGCTTTAAAGCTGCTGACAACGAC	2633
Qy	769	ProThrAlaProProTyrAspThrLeuLeuValPheAspTyrGluGlySerGlySerAsp	788
Db	2634	CCCACGGCGCACCGTATGATCTCCTCTTAGTCTTTGACTACGAGGCGACGGCTCCACG	2693
Qy	789	AlaAlaSerLeuSerSerLeuThrSerSerAlaSerAspGlnAspGlnAspTyrAspTyr	808
Db	2694	GCTGGCTCTTGTAGTCTCCTCACTCCTCAGTAGCGGTGGGACCCAGGACTATGACTAC	2753
Qy	809	LeuAsnGluTrpGlySerArgPheLysLysLeuAlaAspMetTyrGlyGlyGluAsp	828
Db	2754	CTGAATGACTGGGGACCCCGCTTCAAGAAACTGCGGACATGTACGCGCGTGTGTACGAC	2813
RESULT 6			
BM923882			
LOCUS	BM923882	1016 bp	mRNA
DEFINITION	AGENCOURT_6709933 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5760016		
	5', mRNA sequence.		
ACCESSION	BM923882		
VERSION	BM923882.1	GI:19374261	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 1016)		

/note=Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
Subtracted with brain, liver, lung, kidney and muscle.
Directionally cloned. Priming method: oligo-dT. Average
insert size: 1800 bp. Library amplification: 26,000 fold.
Kristi A. Eglund, James J. Vincent, Robert Strausberg,
Bungkook Lee & Ira Pastan: Discovery of new breast
cancer genes encoding membrane and secreted proteins.
Manuscript submitted."

ORIGIN

Alignment Scores:
Pred. No.: 4,88e-144 Length: 873
Score: 1459.00 Matches: 283
Percent Similarity: 97.26% Conservative: 1
Best Local Similarity: 96.92% Mismatches: 5
Query Match: 33.39% Indels: 3
DB: 14 Gaps: 0

US-09-916-849A-1 (1-829) x CA453911 (1-873)

Qy 452 GlyIleProThrGlyGluProValCysValThrAlaGluAspProAspLysGluAsn 471
Db 3 GGGATCCCACTGGGAGCCCTGTGTCTACACTGCAGAGACCCCTGCACAGGAGAA 62

Qy 472 GlnLysIleSerTyrArgIleLeuArgAspProHlaGlyTrpLeuAlaMetAspProAsp 491
Db 63 CAAAGATACAGTACCGCATCTCGAGAGACCCAGCGGTGGCTAGCCATGGACCCAGAC 122

Qy 492 SerGlyGlnValThrAlaValGlyThrLeuAspArgGluAspGluGlnPheValArgAsn 511
Db 123 AGTGGCAGGTTCACAGCTGTGGCACCCCTGCACCGTGAGGATGAGCACTTTGTGAGGAAC 182

Qy 512 AsnIleTyrGluValMetValLeuAlaMetAspAsnGlySerProThrThrGlyThr 531
Db 183 AACATCTATGAAGTCATGGTCTTGGCCATGGACAAATGAAGCCCTCCACCACTGGCAGC 242

Qy 532 GlyThrLeuLeuLeuThrLeuIleAspValAsnAspHisGlyProValProGluProArg 551
Db 243 GGAACCTCTCTGTAACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 302

Qy 552 GlnIleThrLysCysAsnGlnSerProValArgHisValLeuAsnIleThrAspLysAsp 571
Db 303 CAGATCACCATCTGCACCAAGAGCCCTGTGGCCAGGTGCTGAACATCAGGCAAGGAC 362

Qy 572 LeuSerProHisThrSerProPheGlnAlaGlnLeuThrAspSerAspIleTyrTrp 591
Db 363 CTGTCTCCCACTCCCTCTTCCAGCCAGCTCACATGACTGACTGACTGACTGACTGACTG 422

Qy 592 ThrAlaGluValAsnGluGluGlyAspThrValValLeuSerLeuLysLysPheLeuLys 611
Db 423 ACGGCAGAGGTCAACGAGGAGGTGACACAGTGGTCTGTCCCTGAGAGATTCCTGAAG 482

Qy 612 GlnAspThrTyrAspValHisLeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThr 631
Db 483 CAGGATACATATGACGTGACCTTCTCTCTGTGACCATGGCAACAAAGAGAGCTGACG 542

Qy 632 ValIleArgAlaThrValCysAspCysHisGlyHisValGluThrCysProGlyProTrp 651
Db 543 GTGATCAGGCCCACTGTGCGACTGCGCATGGCCATGTCGAAACCTGCGCTGAGCCCTGG 602

Qy 652 LysGlyGlyPheLeuLeuProValLeuGlyAlaValLeuAlaLeuPheLeuLeuLeu 671
Db 603 AAAGGAGGTTTCATCTCTCTCTGCTGGGGGCTGTCTGCTGTCTCTCTCTCTCTCTG 662

Qy 672 ValIleLeuLeuLeuValArgLysLysArgLysLysLysGluProLeuLeuLeuProGlu 691
Db 663 GTCTGCTTTTGTGGTGAAGAGAGCGGAAGATCAAGAGAGCCCTCTCTCTCTCTCTCT 722

Qy 692 AspAspThrArgAspAsnValPheTyrTyrGlyGlu-GluGlyGlyGlyGlyGlyGly 711
Db 723 GATGACACCGGTGACACGCTCTCTACTATGGCGAAAGAGGGGGGTGGCGAGAGACCA 782

Qy 711 nAspTyrAspIleThrGlnLeuHisArgGlyLeuGlu-AlaArgProGluValValLeu- 730

Db 783 GGACTATGATCACCACCAGCTCCACCAAGTGTGGAGCCAGCCCGAGGTGTTCTCC 842

Qy 731 ArgAspAspValAlaProThrIleLeuPro 740

Db 843 CGCAATGAGCTGGCACCACCAACATCATCCG 872

RESULT 8

CA455046

LOCUS

DEFINITION

CA455046 887 bp mRNA linear EST 12-NOV-2002

AGENCOURT_10714428 MAPcL Homo sapiens cDNA clone IMAGE:6722552 5';

CA455046

ACCESSION

VERSION

CA455046.1 GI:24905371

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 887)

AUTHORS

NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgpbs-remail.nih.gov
Tissue Procurement: Kristi A. Eglund, Ira Pastan
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM14285 row: h column: 08
High quality sequence stop: 594.

FEATURES

1..887

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6722552"

/cell_lines="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
hTERT-HME1, LNCap"

/lab_host="EMDH103"

/clone_lib="MAPcL"

/note=Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
Subtracted with brain, liver, lung, kidney and muscle.
Directionally cloned. Priming method: oligo-dT. Average
insert size: 1800 bp. Library amplification: 26,000 fold.
Kristi A. Eglund, James J. Vincent, Robert Strausberg,
Bungkook Lee & Ira Pastan: Discovery of new breast
cancer genes encoding membrane and secreted proteins.
Manuscript submitted."

ORIGIN

Alignment Scores:
Pred. No.: 5,29e-142 Length: 887
Score: 1440.00 Matches: 284
Percent Similarity: 96.97% Conservative: 4
Best Local Similarity: 95.62% Mismatches: 5
Query Match: 32.96% Indels: 4
DB: 14 Gaps: 0

US-09-916-849A-1 (1-829) x CA455046 (1-887)

Qy 287 AspArgGluLysValProGluTyrThrLeuThrIleGlnAlaThrAspMetAspGlyAsp 306

Db 3 GACCCGAAAAAGTCCCTGAGTACACTGACCTCCAGGCCAGACATGATGGGAC 62

Qy 307 GlySerThrThrAlaValAlaValValGluIleLeuAspAlaAsnAspAlaPro 326

Db 63 GGTCTCCACCAACCGAGTGGAGTGTGAGATCTTGTATGCCATGACATGCTCTCC 122

Qy 327 MetPheAspProGlnLysTyrGluAlaHisValProGluAsnAlaValGlyHisGluVal 346

123 ATGTTTACCCCAAGAGTACGAGCCATGTGCTGAGAACTGAGTGGCCATGAGGTG 182
 QY GlnArgLeuThrValThrAspLeuAspAlaProAsnSerProAlaThrArgAlaThrTyr 366
 Db CAGAGGCTGAGGTCCTGATCTGAGCGCCCACTCAGAGGTGGCTGCGACCTAC 242
 QY LeuIleMetGlyLysAspAspGlyAspHisPheThrIleThrHisProGluSerAsn 386
 Db CTTATCATGCGCGTGCAGCGGACCATTTTACCATCACCACCCACCTGAGAGCAAC 302
 QY GlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAlaLysAsnGlnHisThrLeu 406
 Db CAGGGCATCTGACACACAGGAGGTTTGGATTTCAGGCCCAAAACACGACACCCCTG 362
 QY TyrValGluValThrAsnGluAlaProPheValLeuLysLeuProThrSerThrAlaThr 426
 Db TACGTTGAAGTGACCAACAGGAGCCCTTTTGTGCTGAAGCTCCCAACCTCCAGCCACC 422
 QY IleValValHisValGluAspValAsnGluAlaProValPheValProProSerLysVal 446
 Db ATAGTGTCCACGTGAGAGATGTGAATGAGGCACCTGTGTTGTCACCCCTCCAAAGTC 482
 QY ValGluValGlnGluGlyIleProThrGlyGluProValCysValTyrThrAlaGluAsp 466
 Db GTTGAGGTCCAGGAGGATCCCACTGGGAGGCTGTGTGTCTACACTGCAGAGAAC 542
 QY ProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArgAspProAlaGlyTyrLeu 486
 Db CCTGACAGAGAGATCAAAAGATCAGCTACCGCATCTGAGAGACCAGGAGGTGGCTA 602
 QY AlaMetAspProAspSerGlyGlnValThrAlaValGlyThrLeuAspArgGluAspGlu 506
 Db GCCATGGACCCAGACAGTGGGAGGTCACAGCTGTGGGACCCCTCGACCTGAGGATGAG 662
 QY GlnPheValArgAsnAsnIleTyrGluValMetValLeuAlaMetAspAsnGlySerPro 526
 Db CAGTTTGTGAGGAAACACATCTATGAAGTCACTGTGCTGTGGCCATGGACAATGGAAGCCCT 722
 QY ProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAspValAsnAspHisGlyPro 546
 Db CCCACCATGACGGGAAACCTTCTGCTACACTGATGATGATGATGATGATGATGATGATG 782
 QY ValProGlu-ProArgGlnIleThrIleCysAsnGlnSerProVal-ArgHisValLeuA 566
 Db GTCCCTGAGCCCGCTCAGATCAGCATCTGCAACCAAGCCCTGTGCGCCAGGTGCTTG 842
 QY snIleThrAsp-LysAspLeu-SerProHisThrSerProPhe 579
 Db ACATCAGGACAAGGAAGTGTGTTCCCAAACTTCCCTTTT 885

RESULT 9
 BQ936686
 LOCUS
 DEFINITION AGENCOURT_10035567 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6479780
 5', mRNA sequence.
 BQ936686
 VERSION BQ936686.1 GI:22352069
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1049)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 CONTACT Robert Strausberg, Ph.D.
 EMAIL cgapbs-remail.nih.gov
 Tissue Procurement: DCTD/DP
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Plate: LNCM258 row: h column: 21
 High quality sequence stop: 609.
 Location/Qualifiers
 1..1049
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6479780"
 /tissue_type="carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_40"
 /note="Organ: prostate; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.75e-139 Length: 1049
 Score: 1417.50 Matches: 292
 Percent Similarity: 91.44% Conservative: 7
 Best Local Similarity: 89.30% Mismatches: 28
 Query Match: 32.44% Indels: 6
 DB: 13 Gaps: 0
 US-09-916-849A-1 (1-829) x BQ936686 (1-1049)

QY 452 GlyIleProThrGlyGluProValCysValTyrThrAlaGluAspProAspLysGluAsn 471
 Db 3 GGCATCCCACTGGGAGGCTGTGTGTCTACACTGCAGAGACCCCTGACAGGAGAAAT 62
 QY 472 GlnLysIleSerTyrArgIleLeuArgAspProAlaGlyTyrLeuAlaMetAspProAsp 491
 Db 63 CAAAGATCAGCTACCGCATCTGAGAGCCAGCAGGAGGTGGCTAGCCATGACCCAGAC 122
 QY 492 SerGlyGlnValThrAlaValGlyThrLeuAspArgGluAspGluGlnPheValArgAsn 511
 Db 123 AGTGGCAGGTCAAGCTGTGGCACCTCGACCGTGAGGATGAGCACTTTGTGAGGAAC 182
 QY 512 AsnIleTyrGluValMetValLeuAlaMetAspAsnGlySerProProThrThrGlyThr 531
 Db 183 AACATCTATGAAGTCATGCTTGGCCCATGGACAATGGAAGCCCTCCACCATGGCAGC 242
 QY 532 GlyThrLeuLeuLeuThrLeuIleAspValAsnAspHisGlyProValProGluProArg 551
 Db 243 GGAACCTTCTGCTAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 302
 QY 552 GlnIleThrIleCysAsnGlnSerProValArgHisValLeuAsnIleThrAspLysAsp 571
 Db 303 CAGATCACCATCTGCAACCAAGCCCTGTGCGCCAGGTGCTGAACATCAGCGACAGGAC 362
 QY 572 LeuSerProHisThrSerProPheGlnAlaGlnLeuThrAspAspSerAspIleTyrTrp 591
 Db 363 CTGTCTCCCAACACCTTCCCTTTCAGGCCCAAGCTCAGATGATGATGATGATGATGATG 422
 QY 592 ThrAlaGluValAsnGluGlyAspThrValValLeuSerLeuLysGlyPheLeuLys 611
 Db 423 ACGGACAGGTCAACGAGGAGGTGACACAGTGTGCTTGTCTGCTGCTGCTGCTGCTGCTG 482
 QY 612 GlnAspThrTyrAspValHisLeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThr 631
 Db 483 CAGGATACATATGAGTGCACCTTCTCTGTCTGACCATGGCAACAGAGAGCTGACG 542
 QY 632 ValIleArgAlaThrValCysAspCysHisGlyHisValGluThrCysProGlyProTrp 651
 Db 543 GTGATCAGGCGCCACCTGTGTGCGACCTGCCATGCGCAATGTCGAAACCTGCTGCTGCTG 602


```

5', mRNA sequence.
ACCESSION BQ690885
VERSION BQ690885.1 GI:21816201
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
1 (bases 1 to 846)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LUCM2360 row: f column: 15
High quality sequence stop: 599.
FEATURES
source
1..846
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6206222"
/tissue type="ductal carcinoma, cell line"
/lab host="DH10B (phage-resistant)"
/clone lib="NIH_MGC_110"
/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming. The
directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
ORIGIN
Alignment Scores:
Pred. No.: 6.45e-136 Length: 846
Score: 1382.50 Matches: 267
Percent Similarity: 96.40% Conservative: 1
Best Local Similarity: 96.04% Mismatches: 8
Query Match: 31.64% Indels: 2
DB: 13 Gaps: 1
US-09-916-849A-1 (1-829) x BQ690885 (1-846)
QY 553 ILeThrIleCysAsnGlnSerProValArgHisValLeuAsnIleThrAspLysAspLeu 572
Db 2 ATCCACATCTGCACCAAGCCCTGTGGCCACGTGTGACATCAGGACAGGACCTG 61
QY 573 SerProHisThrSerProPheGlnAlaGlnLeuThrAspSerAspIleThrTpThr 592
Db 62 TCTCCCAACACCTCCCTTTCCAGGCCAGCTCACAGATGACTCAGACATCTACTGGACG 121
QY 593 AlaGlnValLeuGlnGluGlyAspThrValValLeuSerLeuLysLysPheLeuLysGln 612
Db 122 GCAGAGTCAACGAGGAGGAGGACACAGTGTCTTGTCCCTGAAGAGTTCCTGAGACAG 181
QY 613 AspThrTyrAspValHisLeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrVal 632
Db 182 GATACATATGACGTGCACCTTCTCTGTCTGACCATGGCAACAAGAGCAGTGCAGGTG 241
QY 633 IleArgIleThrValCysAspCysHisGlyHisValGlnThrCysProGlyProThrLys 652
Db 242 ATCAGGGCCACTGTGTGCGACTGCATGGCCATGTGAAACCTGCCCTGGACCTGGAA 301
QY 653 GlyGlyPheIleLeuProValLeuGlyAlaValLeuAlaLeuPheLeuLeuVal 672

```

```

Db 302 GGAGGTTTCATCTCCCTGTGCTGGGGCTGTCTGGGCTCTGCTCTCTCTCTGCTGTG 361
QY 673 LeuLeuLeuValArgLysLysArgLysLysArgLysLysGluProLeuLeuLeuProGluAsp 692
Db 362 CTGCTTTTGTGTGAGAAAGAGAGCGGAGATCAGAGAGCCCTCTACTATCCAGAGAT 421
QY 693 AspThrArgAspAsnValPheTyrTyrGlyGluGluGlyGlyGlyGluAspGlnAsp 712
Db 422 GACACCCGTGACACGCTCTTCTACTATATGCGGAAGAGGGGGTGGCGAAGAGGACCGAC 481
QY 713 TyrAspIleThrGlnLeuHisArgGlyLeuGluAlaArgProGluValValLeuArgAsn 732
Db 482 TATGACATCACCAGCTCCAGAGTCTGAGGCGCAGGCGGAGGTGTCTCTCGCAT 541
QY 733 AspValAlaProThrIleProThrProMetTyrArgProArgProAlaAsnProAsp 752
Db 542 GACGTGGCACCACCAACCATCATCCGACACCCATGTACCGTCTAGGCCAGCCACCCAGAT 601
QY 753 GluIleGlyAsnPheIleGluAsnLeuLysAlaAlaAsnThrAspProThrAlaPro 772
Db 602 GANATCGGCACCTTTATTAATTGAGAACCTTGAGGCGGCTAACACAGACCCACAGCCCCG 661
QY 773 ProTyrAspThrLeuLeuValPheAspTyrGluGlySerGlySerAspAlaAsnLeu 792
Db 662 CCTACGACACCCCTCTGTGTGCTGCTGAGGCGCAGCGCTCCGACCGCGCTCCCTG 721
QY 793 SerSerLeuThrSerSerAlaSerAspGlnAspGlnAspTyrAspTyrLeuAsnGluTrp 812
Db 722 AGCTCCCTCACCCTCCCTCCGCTCGCCTCGACCAAGACCAAGATTACGATTATCTGAACGAGTGG 781
QY 813 GlySerArg-PheLysLysLeuAlaAspMet---TyrGlyGlyGlyGluAsp 828
Db 782 GGCACGCCCTTCAGAGAGCTGGCAGACATGTAAAGTGGCGGGGGAGGAC 833
CA454560 918 bp mRNA linear EST 12-NOV-2002
AGENCOURT 10714108 MAPCL Homo sapiens cDNA clone IMAGE:6720636 5',
mRNA sequence.
CA454560.1 GI:24904418
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
1 (bases 1 to 918)
REFERENCE National Institutes of Health, Mammalian Gene Collection (MGC)
AUTHORS Unpublished (1999)
TITLE Contact: Robert Strausberg, Ph.D.
JOURNAL Email: cgapbs-remail.nih.gov
COMMENT Tissue Procurement: Kristi A. Eglund, Ira Pastan
CDNA Library Preparation: Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14280 row: h column: 12
High quality sequence stop: 650.
FEATURES
source
1..918
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6720636"
/cell line="2R-75-1, MCF7, SK-BR-3, MDA-MB-231,
hTERT-HME1, LNCaP"
/lab host="EMDH10B"
/clone lib="MAPCL"
/note="Vector: pCMV-SPORT6; Site_1: EcoRV; Site_2: Not I;

```

Subtracted with brain, liver, lung, kidney and muscle.
 Directionally cloned. Priming method: oligo-dT. Average
 insert size: 1800 bp. Library amplification: 26,000 fold.
 Kristi A. Egland, James J. Vincent, Robert Strausberg,
 Bungkok Lee & Ira Pastan: Discovery of new breast
 cancer genes encoding membrane and secreted proteins.
 Manuscript submitted."

ORIGIN

Alignment Scores:

Pred. No.: 7,42e-136 Length: 918
 Score: 1382.50 Matches: 284
 Percent Similarity: 95.03% Conservative: 3
 Best Local Similarity: 94.04% Mismatches: 7
 Query Match: 31.64% Indels: 8
 DB: 14 Gaps: 3

US-09-916-849A-1 (1-829) x CA454560 (1-918)

QY 180 LysTyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMet 139
 DB 3 AAGTATGAGCTCTTTGGCCACGCTGTGTACAGATGTCCTCAGTGGAGACCCCATG 62
 QY 200 AsnIleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThr 219
 DB 63 AACATCTCCATCATGTCGACCGACCAATGACCAAGCCCAAGTTTACCCAGGACAC 122
 QY 220 PheArgLysSerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAla 239
 DB 123 TTCGAGGGAGTGCTTTAGAGGGAGTCTCTACAGGTACTTCTGTGATGACAGGTGACAGCC 182
 QY 240 ThrAspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSer 259
 DB 183 ACGATGAGGATGATGCCATCTACACCTACATGAGGGGTGTGCTTACTCTCATCAGTAC 242
 QY 260 GlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIle 279
 DB 243 CAAGAACCAAGGACCCACACGACCTCATGTTCCACCATTCACCGAGCACAGGACCATC 302
 QY 280 SerValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGln 299
 DB 303 AGCTGTATCTCCAGTGCCTTGGACCGGAAAGTCTCTGATACACACGACGACATCCAG 362
 QY 300 AlaThrAspMetAspGlyAspGlySerThrThrAlaValAlaValAlaValGluIleLeu 319
 DB 363 GCCACACATGATGGGACGGCTCCACACACGCGAGTGGCAGTAGTGAGATCCTT 422
 QY 320 AspAlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGlu 339
 DB 423 GATGCCAATGACATGCTCCATGTTTGACCCCGAGAGTACGAGGCCCATGTGCTGTAG 482
 QY 340 AsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSer 359
 DB 483 AATGACATGGCCATGAGGTGCAGAGGCTGACGGTCACTGATCTGACGCCCCCACTCA 542
 QY 360 ProAlaTrpArgAlaThrTyrLeuIleMetGlyAspAspGlyAspHisPheThrIle 379
 DB 543 CCAGCGTGGCGTGCCACCTACCTTATCATGGCGGTGACGACGCGGACCATTTTACCATC 602
 QY 380 ThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArg-LysGlyLeuAspPheG1 399
 DB 603 ACCACCCACCTGAGACGACACGAGGATCTGTGACACCGAGGAGGGTTTGGATTGTA 662
 QY 399 uAlaLysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeu 419
 DB 663 GGCCAAAACACGACACCTGTACGTTGAAGTGACCAACGAGGCCCTTTTGTGTGAA 722
 QY 419 sLeuProThrSerThrAlaThrIleValValHis-ValGluAspValAsnGluAlaPro 439
 DB 723 GCTCCCAACCTCCACAGCCACCATAGTGGTCCAGGTGGAGGATGTGAATGGAGCACCTG 782
 QY 439 alPheValPro-ProSerLysValVal-GluValGln-GluGlyIlePro---ThrGlyG 457

DB 783 TGTGTGTCCACCCCTCCANAGTCGTGTGAGGTCGAGGAGGCGCATTCCTCCCTCGGGGA 842
 QY 457 luProVal---CysValTyrThrAlaGluAspPro---AspLysGluAsnGlnLys 473
 DB 843 AGCTGTGTGTGTCTACACCTGGCAGAGACCCCTGGCAAGGAAGATTCAAA 858

RESULT 13

BQ928283

LOCUS DEFINITION BQ928283 912 bp mRNA linear EST 20-AUG-2002
 5', mRNA sequence.

ACCESSION BQ928283

VERSION BQ928283.1 GI:22343314

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 912)

AUTHORS

NIH-MGC <http://mgi.nci.nih.gov/>

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: DCTD/DTF
 cDNA Library Preparation: Rubin Laboratory
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM2663 row: d column: 08
 High quality sequence stop: 585.

FEATURES

source

1..912
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6481591"
 /tissue_type="carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_40"
 /note="Organ: prostate; Vector: pOTB7; Site: 1: XhoI;
 Site 2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 1,53e-135 Length: 912
 Score: 1379.50 Matches: 282
 Percent Similarity: 93.11% Conservative: 2
 Best Local Similarity: 92.46% Mismatches: 11
 Query Match: 31.57% Indels: 10
 DB: 13 Gaps: 2

US-09-916-849A-1 (1-829) x BQ928283 (1-912)

QY 452 GlyIleProThrGlyGluProValCysValTyrThrAlaGluAspProAspLysGluAsn 471
 DB 3 GGCATCCCCACCTGGGGAGCCTGTGTGTCTACCTGCAGAGACCCCTGCACAGAGAT 62
 QY 472 GlnLysIleSerTyrArgIleLeuArgAspProAlaGlyTyrLeuAlaMetAspProAsp 491
 DB 63 CAAAGATCAGTACCGCATCTCTGAGACCCAGGAGGGTGGCTAGCCATGACCCAGAC 122
 QY 492 SerGlyGlnValThrAlaValGlyThrLeuAspArgGluAspGluGlnPheValArgAsn 511
 DB 123 AGTGGGAGGTACACAGTGTGGGCCACCTCGACCGTGAGGATGACAGTGTGTGAGGAAC 182

QY 512 AsnIleTyrGluValMetValLeuAlaMetAspAsnGlySerProThrThrGlyThr 531
 Db 183 AACATCATGAAGTCATGGTCTTGGCCATGGACATGAAGCCCTCCACCACCTGGCAGC 242
 QY 532 GlyThrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 551
 Db 243 GGAACCCCTTCGCTAACACTGATGATGATCAATGACCATGCCAGTCCCTGAGCCCGT 302
 QY 552 GlnIleThrIleCysAsnGlnSerProValArgHisValLeuAenIleThrAspLysAsp 571
 Db 303 CAGATCACCATCTGACACCAAGCCCTGTGGCCAGGTGTGAACATCAGGACAGGAC 362
 QY 572 LeuSerProHisThrSerProPheGlnAlaGlnLeuThrAspAspSerAspIleTyrTIP 591
 Db 363 CTGTCTCCCCACAGCTCCCTTTCCAGGCCCCAGCTCAGATGACTCAGACATCTACTGG 422
 QY 592 ThrAlaGluValAsnGluGluGlyAspThrValValLeuSerLeuLysLysPheLeuLys 611
 Db 423 ACGCAGAGGTCAACGAGGAAGTGCACAGTGTCTGTGTCCTGAGAAAGTTCCTGAAG 482
 QY 612 GlnAspThrTyrAspValHisLeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThr 631
 Db 483 CAGGATACATATGACGTGCACCTTCTGTCTGACCATGGCAACAAAGAGCAGCTGACG 542
 QY 632 ValIleArgAlaThrValCysAspCysHisGlyHisValGluThrCysProGlyProTIP 651
 Db 543 GTGATCAGGGCCACTGTGTGCACTGGCCATGCGCATGTGCAAACTGCCCTGACCCCTGG 602
 QY 652 LysGlyGlyPheIleLeuProValLeuGlyAlaValLeuAlaLeuLeuPheLeuLeuLeu 671
 Db 603 AAGGAGGTTTCATCTCCCTGCTGCTGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 662
 QY 672 ValLeuLeuLeuLeuValArgLysArgLysIleLysGluProLeuLeuLeuProGlu 691
 Db 663 GTGCTGCTTTTGTGGTGAAGAAAGCGGAAGATCAAGAGGCGCCCTCTACTCCCGAA 722
 QY 692 AspAspThr-ArgAspAsnValPheTyrTyrGlyGluGluGlyGlyGlyGlyGlyGly 711
 Db 723 GATGACACCCCGTGACACAGCTCTTACTATGGCGAAAGGGGTGGCGGAAAGAGACC 782
 QY 711 In-AspTyrAspIleThr-GlnLeuHisArg-GlyLeuGlu-AlaArgPro----GluVal 728
 Db 783 AGGACTATGACATCACCCAGCTCCACCGAGGTCTGGAAGGCCAGGCCCGGGAAGTGGG 842
 QY 729 ValLeuArg-AsnAspValAla-----ProThrIleIleProThrProMetTyrArgPr 746
 Db 843 TTCCTCCGAAATGACGTGGGACCCCAACCATCATCTTCCCGAAACCATGGTACGGTCC 902
 QY 746 o 746
 Db 903 C 903

RESULT 14
 LOCUS CA489015
 DEFINITION AGENCOURT_10809410 MAPCL Homo sapiens cDNA clone IMAGE:6721249 5',
 mRNA sequence.
 ACCESSION CA489015
 VERSION CA489015.1 GI:24951806
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 875)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Kristi A. Eglund, Ira Pastan
 CDNA Library Preparation: Invitrogen Corp

cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM14282 row: b column: 01

High quality sequence stop: 647.

FEATURES

Location/Qualifiers

1..875

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:6721249"

/cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,

hTERT-HME1, LNCap"

/lab_host="EMDH10B"

/clone_lib="MAPCL"

/note="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;

Subtracted with brain, liver, lung, kidney and muscle.

Directionally cloned. Priming method: oligo-dT. Average

insert size: 1800 bp. Library amplification: 36,000 fold.

Kristi A. Eglund, James J. Vincent, Robert Strausberg,

Bungkook Lee & Ira Pastan: Discovery of new breast

cancer genes encoding membrane and secreted proteins.

Manuscript submitted."

ORIGIN

Alignment Scores:

Pred. No.: 1,976-132 Length: 875

Score: 1350.00 Matches: 262

Percent Similarity: 98.87% Conservative: 1

Best Local Similarity: 98.50% Mismatches: 2

Query Match: 30.90% Indels: 2

DB: 14 Gaps: 0

US-09-916-849A-1 (1-829) x CA489015 (1-875)

QY 565 LeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGlnAlaGlnLeuThr 584
 Db 3 CTGAACATCAGGACAGGACCTGTCTCCCC-ACCTCCCTTTCCAGGCCAGCTCACA 61
 QY 585 AspAspSerAspIleTyrTIPThrAlaGluValAsnGluGlyAspThrValValLeu 604
 Db 62 GATGACTCAGACATCTACTGACGCGCAGGTCAACAGGAAGGTGACACAGTGTCTTG 121
 QY 605 SerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSerLeuSerAspHis 624
 Db 122 TCCTCGAAGAGTTCCTCGACGAGATACATATGACGTGACCTTCTCTGTGACCAT 181
 QY 625 GlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCysHisGlyHisVal 644
 Db 182 GGCAACAAAGAGCAGCTGACGTGATCAGGGCCACTGTGTGCGACTGCCATGGCCATGTC 241
 QY 645 GluThrCysProGlyProTIPLysGlyGlyPheIleLeuProValLeuGlyValValLeu 664
 Db 242 GAAACTGCTCCCTGACCTCGNAGGAGGTTCATCTCCCTGTGTGCTGGGGCTGTCTGT 301
 QY 665 AlaLeuLeuPheLeuLeuLeuValLeuLeuLeuValArgLysLysArgLysIleLys 684
 Db 302 GCTCTGCTGTTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 361
 QY 685 GluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyrTyrGlyGluGlu 704
 Db 362 GAGCCCTCTCTACTCCAGAGATGACCCGCTGACACAGTCTTCTACTATGGCGAGAG 421
 QY 705 GlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArgGlyLeuGluAla 724
 Db 422 GGGGGTGGGGAAGAGCAGGACTATGACATCACAGCTCCACAGCTCCACCGAGGTCTGGAGGCC 481
 QY 725 ArgProGluValValLeuArgAsnAspValAlaProThrIleIleProThrProMetTyr 744
 Db 482 AGGCCGGAGGTGTTCTCCGCAATGACGTGGACCAACCAATCATCCCGACACCCATGTAC 541

```

QY 745 ArgProA:ProAlaAsnProAspGluIleGlyAsnPhelIleGluAsnLeuIysAla 764
Db 542 CGTCTCGCCGAGCCCAACCCAGATGAAATCGGCACTTTATTAATTGAGAACTCGAGGCG 601
QY 765 AlaAsnThrAspProThrAlaProProTyrAspThrLeuValPheAspTyrGluGly 784
Db 602 GCTAACACAGACCCCAAGCCCGCCCTTACGACACCTCTTGGTGTTCGATATGAGGCG 661
QY 785 SerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSerAspGln 804
Db 662 AGCGGCTCCGAGCGCGCTGCTGAGCTCCCTCACCTCTCCGCTCCGACCAAGACCCA 721
QY 805 AspTyrAspTyrLeuAsnGluTyrGlySerArgPheIysLysLeuAlaAspMetTyr-GI 824
Db 722 GATTACGATTATCTGAACAGAGTGGCGAGCGCGCTTCAAGAGCTGGCAGACATGTACCG 781
QY 824 YGlyGlyGluAspAsp 829
Db 782 TGGCGGGGGAAGAGAC 797

RESULT 15
BO676796
LOCUS
DEFINITION BO676796 913 bp mRNA linear EST 15-JUL-2002
5', mRNA sequence.
AGENCOURT 8184299 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6264044
BO676796
VERSION BO676796.1 GI:21789475
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
MAMMALIA; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 913)
NH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2428 row: o column: 21
High quality sequence stop: 635.
FEATURES
source
1..913
location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6264044"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_112"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 1.48e-131 Length: 913
Score: 1343.00 Matches: 277
Percent Similarity: 94.93% Conservative: 4
Best Local Similarity: 93.58% Mismatches: 6
Query Match: 30.74% Indels: 9
DB: 13 Gaps: 2

```

Search completed: September 23, 2004, 01:24:50
Job time : 5141 secs

US-09-916-849a-1 (1-829) x BO676796 (1-913)

```

QY 275 SerThrGlyThrIleSerValIleSerSerGlyLeuAspArgGluLysValProGluTyr 294
Db 2 AGCACAGGACCATCATGAGCGTCATCTCCAGTGGCTGGACCGGAAAGTCCCTGAGTAC 61
QY 295 ThrLeuThrIleGlnAlaThrAspMetAspGlyAspGlySerThrThrAlaValAla 314
Db 62 ACACAGCATCATGAGCGCCACAGCATGGATGGGAGCGGTCCACACACGACGAGTGGCA 121
QY 315 ValValGluIleLeuAspAlaAsnAspAlaProMetPheAspProGlnLysTyrGlu 334
Db 122 GTAGTGGAGATCCTTGATGCCAATGCAATGTCTCCCATGTTTACCCCCAGAGTACGAG 181
QY 335 AlaHisValProGluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeu 354
Db 182 GCCCATGTGCTGAGAAATGAGTGGGCCATGAGGTGCAGAGGTGACGGTCACTGATCTG 241
QY 355 AspAlaProAsnSerProAlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspAspGly 374
Db 242 GACGCCCCCAACTCACAGCGTGGCGTGCACCTACCTTATCATGGCGGTGACGACGGG 301
QY 375 AspHisPheThrIleThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLys 394
Db 302 GACCATTTTACCATCACCCACCCCTGAGAGCAACCCAGGGCATCTGTGACACACCAAGAG 361
QY 395 GlyLeuAspPheGluAlaLysAsnGlnHisThrLeuTyrValGluValThrAsnGluAla 414
Db 362 GGTGTGATTTTGGAGCCCAAAACCCAGCACACCCCTGTAGCTTGAAGTGACCAACGAGGCC 421
QY 415 ProPheValLeuLysLeuProThrSerThrAlaThrIleValValHisValGluAspVal 434
Db 422 CTTTTGTGCTGAAGCTCCCAACCTCCACAGCACCATAGTGTGTCACACGTGGAGGATGTG 481
QY 435 AsnGluAlaProValPheValProProSerLysValValGluValGlnGluGlyIlePro 454
Db 482 ATAGGACACCTGTGTTTGTCCACCCCTCCAAAGTCGTTGAGTCCAGGAGGCGATCCCC 541
QY 455 ThrGlyGluProValCysValTyrThrAlaGluAspProAspLysGluAsnGlnLysIle 474
Db 542 ACTGGGGAGCCTGTGTGTGTCTTACACTGCAGAAAGACCCCTGACAGAGAGAAATCAAAAGATC 601
QY 475 SerTyrArgIleLeuArgAspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGln 494
Db 602 AGTACCGCATCTCTGAGAGACCCAGAGGTGGCTAGCCATGAGCCACGACAGTGGGCGAG 661
QY 495 ValThrAlaValGlyThrLeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyr 514
Db 662 GTCACAGCTGTGGGCACCCCTCGACCGTGAGGATGAGCAGTTTGTGAGGAACACATCTAT 721
QY 515 GluValMetValLeuAlaMetAspAsnGlySerProProThrThrThrGlyThrGlyThrL 534
Db 722 GAAGTCATGGCCCTTGCCCATGGGACATGGAAGCCCTCCACCCACCGGCGGAGAACCC 781
QY 534 euLeuLeuThrLeu-IleAspVal-AsnAsp-HisGly-ProValPro---GluProArg 551
Db 782 TTCTGTAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 841
QY 552 Gln-IleThr-----IleCysAsnGlnSerProVal 561
Db 842 CAGGATCATCCCATCTTGTCCCAACCAAAAGCCCGCTG 875

```

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 22, 2004, 20:44:18 ; Search time 8138 seconds
(without alignments)
4415.259 Million cell updates/sec

Title: US-09-916-849A-1
Perfect score: 4369
Sequence: 1 MGUPRGPLASLLQLQCWLQ.....NEWGSRFKLADMYGGGEDD 829

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09916849/runat_21092004_145803_7009/app_query.fasta_1.967
-DB=GenEmbl -Qfmt=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LISTN=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORW=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09916849@cgn_1_1_5239@runat_21092004_145803_7009 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb_ba.*
2: gb_hg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vi.*
30: em_hig_hum.*
31: em_hig_inv.*
32: em_hig_other.*
33: em_hig_mus.*
34: em_hig_pln.*
35: em_hig_rod.*
36: em_hig_mam.*
37: em_hig_vrt.*
38: em_sy.*
39: em_hgo_hum.*
40: em_hgo_mus.*
41: em_hgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	4369	100.0	3171	6	AR270785	Sequence
2	4369	100.0	3171	6	AX411115	Sequence
3	4369	100.0	3171	6	AX829232	Sequence
4	4369	100.0	3171	9	HSPCAD	X63629 H.sapiens m
5	4358	99.5	3686	9	BC041846	BC041846 Homo sapi
6	4348	99.5	3219	6	AX382118	Sequence
7	3550.5	81.3	3187	10	MMCADHP	X06340 Mouse mRNA
8	3433.5	78.6	2993	10	BC052189	BC052189 Mus muscu
9	2549.5	58.4	2436	5	GGCADHB	X58518 Chicken mRN
10	2420	55.4	4778	6	AX658364	AX658364 Sequence
11	2420	55.4	4778	9	HSECAD	Z13009 H.sapiens m
12	2420	55.4	4828	6	AX587858	Sequence
13	2417	55.3	2649	9	HSECADH	Z18923 H.sapiens m
14	2400.5	54.9	2807	9	AB025106	Sequence
15	2388	54.7	4362	10	YMECADH	AB025106 Homo sapi
16	2388	54.7	4362	10	AB017696	X06115 Mouse mRNA
17	2377	54.4	3450	5	XL004707	AB017696 Rattus no
18	2375	54.4	2699	5	XL004707	U03720 Xenopus lae
19	2371	54.3	3750	5	XL004707	X63720 X.laevius mR
20	2349.5	53.8	2808	6	IS7999	X78546 X.laevius XB
21	2349.5	53.8	2808	6	AR214053	Sequence
22	2349.5	53.8	2808	6	AR369494	Sequence
23	2342.5	53.6	2808	9	HUMVOECAD	L08599 Human uvom
24	2323	52.0	2106	4	BTCDHP	X53614 Bovine mRN
25	2274	50.9	2486	10	MMUVOM	M16260 Chicken liv
26	2223	50.7	2567	9	AB025105	X06339 Mouse mRNA
27	2216.5	50.7	2567	9	AB025105	AB025105 Homo sapi
28	2204.5	50.5	4097	5	XELCADH	L29057 Xenopus lae
29	2198	50.3	2819	5	XL004708	U04708 Xenopus lae
30	2142	49.0	2173	5	XL004708	X75454 X.laevius mR
31	2131	48.8	4434	5	CHKKCAM	M81894 Chicken B-c
32	1893	43.3	3340	5	AF364811	AF364811 Danio rer
33	1707	39.1	2853	9	HSNCAD	X54315 Human mRNA
34	1704.5	39.0	3203	5	GGNCAD	X07277 Chicken mRN
35	1702	39.0	4033	9	BC036470	BC036470 Homo sapi
36	1701.5	38.9	4350	10	AB017695	AB017695 Rattus no
37	1701	38.9	3269	10	BC022107	BC022107 Mus muscu
38	1701	38.9	4321	10	AB008811	AB008811 Mus muscu
39	1699.5	38.9	4132	9	S42303	S42303 N-cadherin
40	1697.5	38.9	4321	6	AX305854	AX305854 Sequence
41	1697.5	38.9	4321	10	MUSCADNA	M31131 Mouse neur
42	1694	38.8	2954	5	CHKRCG	D14459 Gallus gall
43	1693	38.8	3036	10	AF097593	AF097593 Rattus no
44	1691	38.7	3839	4	BTCDHNN	X53615 B.taurus mR
45	1682	38.5	2827	5	AF418565	AF418565 Danio rer

ALIGNMENTS

RESULT 1

[illegible]

```

Db 1794 GCCAGCTCAGAGATGACTCAGACATCTACTGGACGGCAGAGGTCAACGAGGAAGGTGAC 1853
QY ThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGGTCTTGTCCCTGAGAGAGTCTCTGAGCAGAGTACATATGACGTGCACCTTCT 1913
QY LeuSerAspHisGlyAsnLysGluGlnLeuThrValHisArgAlaThrValCysAspCys 640
Db 1914 CTGTCTGACCATGGCAACAAGAGCAGCTGACGGTGTATCAGGGCCACTGTGTGGCACTGC 1973
QY HisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheLeuLeuProValLeu 660
Db 1974 CATGCCATGTGCAAACTGCTGACCTTGGACCTTGGAAAGAGGTTTCACTCCCTGTGCTG 2033
QY GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuLeuValArgLysLys 680
Db 2034 GGGGTGTGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2093
QY ArgLysLeuLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAAGGAGCCCTCTACTCCCAAGAGATGACACCCGTGACACCTCTCTCTAC 2153
QY TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspPheLeuLeuHisArg 720
Db 2154 TATGGCAAGAGGGGGTGGCAAGAGGACCCAGACTATGACATCACCCAGCTCCACCGA 2213
QY GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrLysLysPro 740
Db 2214 GGCTCGAGGCGGCGGAGTGGTCTCCGCAATGACGTGGCACCACCATCATCCCG 2273
QY ThrProMetTyrArgProArgProAlaAsnProAspGluLysGlyAsnPheLysLysGlu 760
Db 2274 ACACCATGTACCTCTAGCCAGCCACCCAGATGAAATCGGCACTTTAATTTAG 2333
QY AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGCGGCTAAACAGACACCCACAGCCCGCCCTAGACACCCCTCTTGGTTC 2393
QY AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATAGGCGACGGCTCCGACGGCGGCTCTGAGCTCCCTGAGCTCCCTCTCCGCTCC 2453
QY AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGACAGGTGGGGCAGCCGCTTCAGAGAGCTGGCA 2513
QY AspMetTyrGlyGlyGlyAspAsp 829
Db 2514 GACATGTACGGTGGCGGGAGGACGAC 2540

RESULT 2
AX411115 3171 bp DNA linear PAT 14-JUN-2002
LOCUS
DEFINITION Sequence 3762 from Patent WO0229103.
ACCESSION AX411115
VERSION AX411115.1 GI:21443820
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 3762 11-APR-2002;
GENE LOGIC INC (US)
FEATURES
source
1. 3171
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon.9606"
/note="EMBL/GenBank Accession No. X63629"

```

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 3171
 Score: 4369.00 Matches: 829
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-916-849A-1 (1-829) x AX411115 (1-3171)

QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuLeuValCysTrpLeuGln 20
 Db 54 ATGGGGCTCCCTCGTGACCTCTCGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113
 QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
 Db 114 TGGCGGGCTCTCGAGCGGTGGCGGGTCTTCAGGAGGCTGAAGTGAAGTGAAGTGA 173
 QY 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
 Db 174 GGAGGGCGGAGCAGGAGCGCGCCAGCGCTGGGAAAAGTATTCTATGGCTGCCCTGG 233
 QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspAspPheThrValArgAsnGlyGluThr 80
 Db 234 CAAGAGCCAGCTCTGTATTAGCACTGATATGATGACTTCACCTGTCGGGAATGGCGAG 293
 QY 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
 Db 294 GTCCAGGAAGAAGGTCACTGAAGGAAGGAATCCATGAAGATCTTCCATCCAAAGCT 353
 QY 101 IleLeuArgArgHisLysArgAspTrpValAlaProIleSerValProGluAsnGly 120
 Db 354 ATCTTACGAAGACACAAGAGAGATTGGGTGGTGTCTCCAAATATCTGTCTCTGAAATGG 413
 QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
 Db 414 AAGGTGCTCTCCCGCAGAGACTGATCAGCTCAAGTCTAATAAGATAGAGACCAAG 473
 QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGlyValPheAlaVal 160
 Db 474 ATTTTCTACAGCATCACGGGGCGGGGCGAGACAGCCCTCTGAGGGTGTCTTCGTGTA 533
 QY 161 GluLysGluThrGlyTrpLeuLeuLeuLysProLeuAspArgGluLysLysLys 180
 Db 534 GAAAGGAGACAGGCTGGTGTGTGTGAATAGCCACTGGACCGGAGAGATTGCCAAG 593
 QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
 Db 594 TATGAGCTCTTTGGCCACGCTGTGTGAGAGAATGTGTCTCAGTGGAGAGCCCATGAAC 653
 QY 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
 Db 654 ATCTCCATCATCGTGACCGACAGAGATGCCAAGCCCAAGTTTACCAGGACACCTTC 713
 QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
 Db 714 CGAGGAGTGTCTTAGAGGAGTCTTACAGGACTCTTCTGTGATGACAGGTGACACCCACA 773
 QY 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
 Db 774 GATGAGATGATGCCATCTACACCTTACCAATGGGTGGTGTCTTACTCCATCCATAGCCAA 833
 QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
 Db 834 GAAACCAAGGACCCACAGCCTTCAATTCACCAATCCCGAGACAGGACCATCAGC 893
 QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
 Db 894 GTCATCTCCAGTGGCTGGACCGGGAAAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
 QY 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValGluIleLeuAsp 320

Db 114 TCGCGGCGCTCGAGCCGTCGGGCGGTCTTCAGGGAGGCTGAAGTACCTTCGAGGCG 173
Qy 41 GlyGlyAlaGluGlnProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db 174 GAGGCGCGGAGCAGGAGCCCGGCGCGCTGGGAAAGTATTTCATGGGCTGCCCTGGG 233
Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
Db 234 CAAGAGCAGCTCTGTTTAGCACCTGATATGATGACTTCACCTGTGCGGAATGGCGAGCA 293
Qy 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
Db 294 GTCCAGGAAAGAGTCACTCAGGAAAGGAATCCATTGAAGATCTCCCATCCAAAGT 353
Qy 101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
Db 354 ATCTTAGCAACACACAGAGAGATGGGTGGTCTCCTCAATATCTGTCCTCGAAAATGGC 413
Qy 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
Db 414 AAGGGTCCCTTCCCCAGAGACTGATCAGCTCAAGTCTAATAAGATAGAGACCAAG 473
Qy 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal 160
Db 474 ATTTTCTACAGCATCAGGGGCGGGGCGACAGCCCTCGAGGGTGTCTTCGCTGTA 533
Qy 161 GluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
Db 534 GAGAGAGAGACAGCTGTTGTTGAATAAGCCACTGGACCGGGAGGAGATTGCCAAG 593
Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 594 TATGAGCTCTTTGGCCACGCTGTGTACAGAAATGGTGCCTCAGTGGAGGACCCCATGAC 653
Qy 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 654 ATCTCCCATCATCGTGCACGACCAAGATGACCAAGCCCAAGTTTACCCAGGACACCTTC 713
Qy 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 714 CGAGGGAGTGTCTAGAGGGAGTCTTACCAAGTACTTCTGTGATGAGGTGACAGCCACA 773
Qy 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
Db 774 GATGAGGATGATGCATCATCACCTACATCAATGGGGTGTGCTTACTCCATCCATAGCAA 833
Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 834 GAACCAAGAGCCACACAGCCTCATGTTCAAAATTCACCGAGCAGGACCATCAGC 893
Qy 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db 894 GTCATCTCCAGTGGCCTGGACCGGAAAAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
Qy 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluIleLeuAsp 320
Db 954 ACAGACATGATGGGACGCTCCACACACCGCAGTGGCAGTAGTGAGATPCTTTGAT 1013
Qy 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1014 GCCAATGCAATGCTCCATGTTTGACCCCCAGAGTACAGGCCCCATGTGCTTGAGAT 1073
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1074 GCAGTGGCCATGAGGTGCAGAGGCTGACCGGTCACTGATCTGACGCCCCCACTCACCA 1133
Qy 361 AlaTrpArgAlaThrTyrLeuIleMetGlyValAspAspGlyAspHisPheThrIleThr 380
Db 1134 GCGTGGCGTCCACCTACCTATATCTGGGGCGTGCAGCGGGACCATTTTACCATCACC 1193
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400

Db 1194 ACCACCCCTGAGAGCAACCCAGGSCATCTCTGACAAACCAGGAAGGTTTGGATTTTGAGGCC 1253
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAAACGACACACCTGTAGTGTGAAGTGACCAACGAGGCCCTTTTGTGCTGAAGCTC 1313
Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCCACCATAGTGTCTCCAGTGGAGGATGTGAATGAGGCACCTGTGT 1373
Qy 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1374 GTCCCACTCCAAAGTCTGTGAGTCCAGAGGGGATCCCACTGGGAGGCTGTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTGCAGAGACCTCACAAGGAGAAATCAAAAGATCAGCTACCGCATCCTGAGA 1493
Qy 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCAGCAGGGTGGCTAGCCATGACCCAGACAGTGGGCAGGTCACTGTGGGCACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGAAGTGAAGTGTGTGAGGAACAACATCTATGAAGTCATGCTTTGGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeuAsp 540
Db 1614 ATGGCAATGGGAAGCCCTCCCACTGGCAGCGGAACCCCTTCTGCTAACACTGATTGAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGGCCAGTCCCTGAGCCCGCTCAGATCACCATCTGCACCAAGACCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGCGCCAGCTGTGAACATCACGACAGGACCTGTCTCCCCACACCTCCCTTTCCAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGluGlyAsp 600
Db 1794 GCCAGCTCACAGATGATCTAGCATCTACTGGCAGCAGAGGTCAAGAGGAAGGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGGTCTTCTCCCTGAAGAAGTCTCTGAAGCAGGATACATATGACGTGCACCTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTGACCATGGCAACAAAGAGCAGCTGACGGTATCAGGGCCACTGTGTGCGACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValLeu 660
Db 1974 CATGCCCATGTGCAAACTGCCCTGGACCTCGAAAGAGGAGTTTCATCTCCCTGTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuPheLeuLeuValLeuLeuLeuLeuValArgLysLys 680
Db 2034 GGGGCTGTCTCGCT 2093
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2094 CGAAGATCAAGAGCCCTCTACTCCAGAGATGACACCCCGTGACAGCTCTTCTAC 2153
Qy 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCAAGAGGGGGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db 2214 GGTCTGGAGGCGAGCGCGAGGTGTTCTCCGCAATGACGTGGGACCAACCATCATCCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db 2274 ACACCATGTACCGTCTCTAGGCCAGCAACCCAGATGAATCGCAACTTTTATAATTGAG 2333

Db	1074	GCAGTGGGCCATGAGTGCAGAGGCTGACGGTCACTGATCTGGACGCCCCCAACTCACCA	1133
QY	361	AlaTTrpArgAlaThrTyrlleuileMetGlyGlyAspAspGlyAspHisPheThrIleThr	380
Db	1134	GGTGGCGGTGCCACCTACTTATCATGTGGCGGTGACGACGGGACCAITTTTACCATCAC	1193
QY	381	ThrHisProGluSerAsnGlnGlyIleuThrThrArgIysGlyLeuAspPheGluAla	400
Db	1194	ACCCACCTGTGAGACACACAGGCACTCTGCAACACGAGAAGGTGTGGATTTTGAGGCC	1253
QY	401	LysAsnGlnHisThrLeuTyrlValGluValThrAsnGluAlaProPheValLeuLysLeu	420
Db	1254	AAAAACACGACACACCTGTAGTTGAAGTGACCAACGAGGCCCTTTTGTGCTGAGCTC	1313
QY	421	ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe	440
Db	1314	CCAACTCTCACAGCCACCATAGTGTCTCACGTGGAGGATGTGAATGAGGCACCTGTGTTT	1373
QY	441	ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys	460
Db	1374	GTCCACACCTCCAAAGTCGTTGAGGTCCAGAGGGCATCCCATCTGGGAGCCCTGTGTGT	1433
QY	461	ValTyThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyArgIleLeuArg	480
Db	1434	GTCTACACTGCAGAGAACCCTGCACAAGAGAAATCAAAAGATCAGTACCGCATCCTCGA	1493
QY	481	AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr	500
Db	1494	GACCCAGCAGGCTGGCTAGCCATGGACCCAGACAGTGGCAGGTCACAGCTGTGGGCACC	1553
QY	501	LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrlGluValMetValLeuAla	520
Db	1554	CTCGACCGTGAAGTATGAGCATTTGTGAGGAACAACATCATGAAGTCATGGTCTTGGCC	1613
QY	521	MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp	540
Db	1614	ATGACAAATGAAGACCCCTCCACACCTGGCAGCGGAAACCCCTCTGCTTAAACACTGAT	1673
QY	541	ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro	560
Db	1674	GTCAACGACCATGCCCCAGTCCCTGAGCCCCGTGAGTACCATCTGCAACCAAGGCCCT	1733
QY	561	ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln	580
Db	1734	GTGGCCACCGTGTGAACATCACGGACAAGGACCTGTCTCCCCACACCTCCCTTCCAG	1793
QY	581	AlaGlnLeuThrAspAspSerAspIleTyrlTriThrAlaGluValAsnGluGlyAsp	600
Db	1794	GCCAGCTCACAGATGACTCAGACATCTATGACGCCAGAGGTCAACGAGGAAGTGAC	1853
QY	601	ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrlAspValHisLeuSer	620
Db	1854	ACAGTGTCTTGTCCCTGAAGAAAGTTCTCTGAAGCAGGATACATATGACGTGCACTTTCT	1913
QY	621	LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys	640
Db	1914	CTGTCTGACCATGGCAACAAGACACTCAGCGTATCAGGCGCACTGTGTGCGACTCG	1973
QY	641	HisGlyHisValGluThrCysProGlyProTriLysGlyGlyPheIleLeuProValLeu	660
Db	1974	CATGGCCATGTGAAACCTCCCTGGACCCCTGGAAAGGAGGTTTCATCCTCCCTGTGCTG	2033
QY	661	GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys	680
Db	2034	GGGGCTGTCTGGCTGTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2093
QY	681	ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr	700
Db	2094	CGGAAGATCAAGGAGCCCTCCTACTCTCCAGAGATGACACCGTGACCAACCTCTCTTAC	2153
QY	701	TyrGlyGluGlyGlyGlyGluAspGlnAspTyrlAspIleThrGlnLeuHisArg	720

Db	2154	TATGGCGAAGAGGGGGTGGCGAAGAGGAGCAGGAGTATGACATCATCCAGCTCCACCGA	2211
Qy	721	GlyLeuGluAlaArgProGluValValLeuAIGAsnAspValAlaProThrIleIlePro	740
Db	2214	GGTCTGGAGCGCCAGCGCGAGGTGGTTCTCCGCAATGACGTGGCACCAACCATCATCCCG	2273
Qy	741	ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu	760
Db	2274	ACACCATGTACCGTCTCTAGCCAGCGCAACCCAGATGAAATCGCACTTTATAATTGAG	2333
Qy	761	AsnLeuIysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe	780
Db	2334	AACCTGAGCGCGCTAACACAGACCCACAGCCCCCGCTACGACCCCTCTTGGGTGTT	2393
Qy	781	AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer	800
Db	2394	GACTATGAGCGCAGCGCTCCGACGCGCGTCTCTGAGTCTCCTACCTCTCCGCGCTCC	2453
Qy	801	AspGlnAsnGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheIysIysLeuAla	820
Db	2454	GACCAAGACCAAGATTACCATTTATCTGACAGAGTGGCGAGCGCTTCAAGAAGCTGGCA	2513
Qy	821	AspMetTyrGlyGlyGlyGluAspAsp	829
Db	2514	GACATGTACGGTGGCGGGAGGACGAC	2540
RESULT 5	BC041846	3686 bp mRNA linear PRI 12-NOV-2003	
LOCUS	BC041846	clone MGC:43574 IMAGE:5569523, complete cds.	
DEFINITION	BC041846	3686 bp mRNA linear PRI 12-NOV-2003	
ACCESSION	BC041846	clone MGC:43574 IMAGE:5569523, complete cds.	
VERSION	BC041846.1	GI:27693123	
KEYWORDS	MGC.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1	(bases 1 to 3686)	
AUTHORS	Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Schmeitz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Igoe,L.N., Peters,G.J., Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.C., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalls,D.E., Schnerch,A., Schein,J.E., Jones,S.J., and Marra,M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	99 (26), 16899-16903 (2002)	
MEDLINE	22388257		
PUBMED	12477932		
REFERENCE	2	(bases 1 to 3686)	
AUTHORS	Strausberg,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-DEC-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk Email: cgapbs-remail.nih.gov Tissue Procurement: Mitkos.Palkovits.M.D.Ph.D		

QY 361 AlaTrpArgAlaThrTyrLeuIleMetGlyAspGlyAspHisPheThrileThr 380
Db 1625 GGGTGGCGTGCACCTACCTATCATGCGCGGTGACGACGGGAGCAATTTACCATCACC 1684
QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1685 ACCACCCCTGAGAGAACAGGGGATCTCTGACCAACAGAGAGGGTTTGATTTTGAGGCC 1744
QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1745 AAAAACCCAGCACACCTGTACGTTGAAGTGACCAAGAGGCCCTTTTGTGCTGAGCTC 1804
QY 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1805 CCAACCTCCACAGCACCATAGTGTCCAGCTGGAGGATGTGAATGAGGCCACCTGTGTGT 1864
QY 441 ValProProSerLysValValGluValGlnGlyIleProThrGlyGluProValCys 460
Db 1865 GTCCCCACCTCCAAAGTGTGTGAGTCCAGAGGGCATCCCACTGGGGGCCCTGTGTGT 1924
QY 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1925 GTCTACACTGAGAGACCTGACAAAGAGAAATCAAAAGATCAGCTACCGCATCTCTGAGA 1984
QY 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1985 GACCAGCAGGGTGTGCTAGCCATGACCCAGACAGTGGCGAGGTACAGCTGTGGGCACC 2044
QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 2045 CTCGACCGTGAAGTGAAGTGTGTGAGGAGCAACATCATGAGTCATGGTCTTGCC 2104
QY 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db 2105 ATGGACAAATGAAGCCCTCCACCACTGGCAGCGGAAACCCCTCTGCTAACACTGATGAT 2164
QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 2165 GTCAACGACCATGCGCCAGTCCCTGAGCCCGTGCAGATCACCATCTGCACCAACGACCT 2224
QY 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 2225 GTGGCGCCAGGTGTGAACATCAGCATCTACTGGACGCGAGGACCTGTCCACACCTCCCTTCCAG 2284
QY 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 2285 GCCCAGCTCAGATGACTCAGACATCTACTGGACGCGAGGTCACAGGAGGAGGTGAC 2344
QY 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 2345 ACAGTGTCTGTGTCCTGGAAGAGTTCCTGAAGCAGGATACATATGACGTGCACCTTCT 2404
QY 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 2405 CTGTCTGACCATGTCACCAAGAGCAGCTACCGTGTATCAGGGCCACTGTGTGCGACTGC 2464
QY 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu 660
Db 2465 CATGGCCATGTGGAACCTGCTCCCTGACCCCTGGAAGAGGTTTATCTCCCTGTGCTG 2524
QY 661 GlyAlaValLeuAlaLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2525 GGGGCTGTCTGCTGCTGTGCTGTCTCTCTCTGCTGCTGCTTGTGTGTGAGAAAGAG 2584
QY 681 ArgLysIleLysGluProLeuLeuLeuProGluAspThrArgAspAsnValPheTyr 700
Db 2585 CGGAAGATCAAGAGGCCCTCTCTACTCCCAAGAGATGACACCCGTGACCAACGCTTCTAC 2644
QY 701 TyrGlyGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2645 TATGGCAAGAGGGGGTGGCAAGAGGACCAAGGACTATGACATCACCAGCTCCACCA 2704

QY 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db 2705 GGTCGTGAGGCCAGCGCGAGGTGTTCTCGCAATCAGCTGGCACCACCATCATCCCG 2764
QY 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db 2765 ACACCCATGTACCGTCTAGGCGAGCCACCCAGATGAAATCGCAACTTTATAATTGAG 2824
QY 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2825 AACCTGAAGCGCGCTAACACAGACCCACAGCCCGCCCTACGCTCCCTCCCTCTTGGTTC 2884
QY 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2885 GACTATGAGGCGAGCGCTCCGACGCGCGTCCCTGAGCTCCCTACCTCTCCGCTCC 2944
QY 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
Db 2945 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCGCTTCAAGAGCTGGCA 3004
QY 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 3005 GACATGTACGTCGCGGGGAGGACGAC 3031
RESULT 6
AX382118 LOCUS 3219 bp DNA linear PAT 18-MAR-2002
DEFINITION Sequence 17 from Patent WO0200939.
ACCESSION AX382118
VERSION AX382118.1 GI:19576927
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Macina,R.A. and Pillai,R.
TITLE Method of diagnosing, monitoring, staging, imaging and treating
JOURNAL colon cancer
Patent: WO 0200939-A 17 03-JAN-2002;
Diadexus, Inc. (US)
FEATURES
Location/Qualifiers
1..3219
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 3219
Score: 4348.00 Matches: 828
Percent Similarity: 99.76% Conservative: 0
Best Local Similarity: 99.76% Mismatches: 1
Query Match: 99.52% Indels: 1
DB: 6 Gaps: 0
US-09-916-849A-1 (1-829) x AX382118 (1-3219)
QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuValCysTrpLeuGln 20
Db 90 ATGGGGCTCCCTCGTGGACCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 149
QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
Db 150 TCCGGCGGCTCCGACCGCTCGCGCGGTCTTCAGGAGGCTGAAGTACCTTGGAGCG 209
QY 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db 210 GGAGGCGCGAGCAGGAGCCCGCAGCGCTGGGGAAGTATTTCATGGCTGCCCTGGG 269
QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
Db 270 CAAGAGCCAGCTCTGTTTGTAGCACTGATGATGACTTCACTGCGGAATGGCGAGACA 329

Qy 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
Db 330 GTCCAGGAAGAGGTCACCTGAAGGAAGGAATCCATTGAAGATCTTCCCATCCAAACGT 389
Qy 101 IleLeuArgArgHisLysArgSerProValValAlaIleProIleSerValProGluAsnGly 120
Db 390 ATCTTACGAGACACAGAGAGATTGGGTGGTTCCTCAATATCTGCTCCCTGAAATGGC 449
Qy 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
Db 450 AAGGTCCTCTCCCCAGAGACTGAATCAGCTCAAGTCTTAATAAGATAGAGACACCAAG 509
Qy 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
Db 510 ATTTTCTACAGCATCACGGGGCGGGGCGAGACGCCCTCCCTGAGGGTGTCTTCGCTGTA 569
Qy 161 GluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
Db 570 GAGAAGGAGACAGCTGGTGTGTGTAATAAGCCACCTGGACCGGAGGAGATTGCCAAG 629
Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 630 TATGAGCTCTTTGGCCACCGCTGTGTCAGAGAATGGTCTCAGTGGAGACCCCATGAAC 689
Qy 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 690 ATCTCCATCATCTGACCCAGCAGATGACCAAGCCCAAGTTTATCCACGACACCTTC 749
Qy 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 750 CGAGGAGAGTCTTATAGAGAGTCTTACCAGGTACTTCTGTGATGACAGGTACAGCCACA 809
Qy 241 AspGluAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
Db 810 GATGAGGATGATGCCATCTACACCTACAAATGAGGGTGGTGTCTTACTCCATCATAGCAA 869
Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 870 GAACCAAGAGACCCACACGACTCATGTTCACAAATCCACGGAGACAGGACCATCAGC 929
Qy 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db 930 GTCATCTCCAGTGGCTGACCGGAGAAAGTCCCTGAGTACACACTGACCATCCAGGCC 989
Qy 301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValValGluIleLeuAsp 320
Db 990 ACAGACATGGATGGGAGCCGCTCCACCCACCGGAGTGGCAGTGGAGATCCTTGAT 1049
Qy 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1050 GCCAATGACATGCTCCCATGTTTACCCCGGAGAGTACGAGGCCCATGTGCTGAGAA 1109
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1110 GCAGTGGGCGATGAGGTGAGAGGCTGACGGTCACTGATCTGGACGGCCCCCACTCACCA 1169
Qy 361 AlaTrpArgAlaThrTyrIleLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db 1170 GGTGGCGTGCACCTTACCTTATATGCGGTGACGACGCGGGACCACTTTTACCATCAC 1229
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1230 ACCACCTTGAGAGAACAGGGCATCTCTGACACAGAGAGGGTGTGGATTTTGAGGCC 1289
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1290 AAAACCCAGCACACCTGTACCTTGAAGTGACCAACGAGGCCCTTTTGTGCTGAAGCTC 1349
Qy 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
Db 1350 CCAACCTCCACAGCCACCATAGTGTCTCCAGCTGGAGGATGTAATGAGGACCTGTGTTT 1409

Qy 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1410 GTCCCAACCTCTCAAAAGTCGTTGAGGTCAGGAGGCGCATCCCACTGGGGAGACCTGTGTGT 1469
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrAsnIleLeuArg 480
Db 1470 GTCTACATGTCAGAGAGACCCCTGACAGAGAGATCAAAAGATCAGCTACCCGATCCTGAGA 1529
Qy 481 AspProAlaGlyTyrTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1530 GACCAGCAGAGGCTGCTACCATGAGCCACACAGTGGCAGGTCACAGCTGTGGGCACC 1589
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1590 CTCGACCTGTAGGATGACAGATTTGTGAGGAAACAATCTATGAGATCATGGTCTTGGCC 1649
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeuAsp 540
Db 1650 ATGGCAATGGAAAGCCCTCCACCACTGGCAGCGGAAACCTTCTGTCTAACACTGATTGAT 1709
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1710 GTCAACGACCATGGCCCGCTCCCTGAGCCCGCTCAGATCACCATCTGCAACCAAGCCCT 1769
Qy 561 ValArgHisVal-LeuAsnIleThrAspLysAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1770 GTGCGCCAGGTGGCTGAACATCAGCGACAGGACCTGTCTCCCCACACCTCCCTTCCA 1829
Qy 580 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGluGlyAs 600
Db 1830 GGCCGAGCTCACAGATGACTCAGACATCTACTGACCGGACAGAGTCAACGAGGAAGTGA 1889
Qy 600 pThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSe 620
Db 1890 CACAGTGGTCTGTCTGCTGAAGAAGTTCCTGAAGCAGGATACATATGACGTGACACCTTC 1949
Qy 620 rIleuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1950 TCTGTCTGACCATGGCAACAAGAGAGAGCTGACGGTGATCAGGGCCACTGTGTGCGACTG 2009
Qy 640 sHisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValle 660
Db 2010 CCATGGCCATGTGCAAACTGCCCTGGACCTGGAAGGAGGTTTCATCTCCCTGTGCT 2069
Qy 660 sGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2070 GGGGCTGTCTGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2129
Qy 680 sArgLysIleLysGluProLeuLeuProGluAspAspThrArgAspAsnValPheTy 700
Db 2130 GCGGAAGATCAAGAGAGCCCTCTCTACTCCAGAGAATGACACCCGTCACACCTCTCTA 2189
Qy 700 rTyrGlyGluGluGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisAr 720
Db 2190 CTATGGCGAAGAGGGGTGGCGAAGAGACAGGACTATGACATCACCCAGCTCCACCG 2249
Qy 720 sGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePr 740
Db 2250 AGGTCTGAGGCCAGGCGGAGGTGTTCTCCGCAATGACGTGGCAGCAACCATCATCC 2309
Qy 740 cThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGln 760
Db 2310 GACACCCATGACCTCTAGGCCAGCAACCCAGATGAATCGGCAACTTATATATGA 2369
Qy 760 uAsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPh 780
Db 2370 GAACCTGAAGCGCTAACACAGACCCCAAGCCCGCTAGCAGACCTCTTGTGTGT 2429
Qy 780 sAspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSe 800
Db 2430 CGACTATGAGGGCAGCGGCTCCGACGCGGCTGCTGAGCTCCCTCCTCCCTCCCTCC 2489
Qy 800 rAspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLysLeuAl 820

Db 2490 CGACCAGACCAAGATTACGATTATCTGAACAGCTGGCGGACCGCTTACAGAGACTGGC 2549

Qy 820 aAspMetTyrGlyGlyGlyGluAsp 829

Db 2550 AGACATGTACGGTGGCGGGAGGACGAC 2577

RESULT 7

MMCADHP

LOCUS

DEFINITION Mouse mRNA for P-cadherin.

ACCESSION X06340

VERSION X06340.1 GI:50267

KEYWORDS cadherin; cell adhesion molecule; glycoprotein; transmembrane protein.

SOURCE

ORGANISM Mus musculus (house mouse)

REFERENCE

AUTHORS Nose.A., Nagafuchi.A. and Takeichi.M.

TITLE Isolation of placental cadherin cDNA: identification of a novel gene family of cell-cell adhesion molecules

JOURNAL EMBO J. 6 (12), 3655-3661 (1987)

MEDLINE 89111554

PUBMED 3428270

FEATURES

source

1..3187

/location=Qualifiers

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/clone="p3"

/cell_line="PSA5-E"

46..2514

/note="unnamed protein product; pre P-cadherin (AA -27 to 793)"

/codon_start=1

/protein_id="CAA29646.1"

/db_xref="GI:50268"

/db_xref="GOA:P10287"

/db_xref="SWISS-PROT:P10287"

/translation="MELLSGPFAFLLLILQVCLRWSEPYRAGPIEAGVTLVEVG
TDLPEQVGLKVALAGCGHHADNGRIIMLTGTGQGGKDMHSPETRIILRKRREW
MPPIFVPEKNGKFFQRLNQLKSDRGTKIFYITPGADSPPEGVFTIEKESGWL
LHMLDREKIVKLYEYCHAVSNGASVEPMNISIIVTDQNDKPKFTODTPRGSVLE
GMPGTSVMQVATDDEDDAVNTGVVAVSIHQEPKPHDLMTTHKSTGTISVSS
GLDRKXVRYLTQVATDMDGEGSTTTAAVQLILDANDNAPEFEPQKKEAVVPEVY
GHEQVRLVTDLDPNMPARWATHIVGGDGDHFTITHPTINQVLTLLKGLDFEA
QOQVLTVEPNEAPFAVLPTATAVTVHVKDNEAPVFPSPKVEIAQEGISIGEL
VCITYAQDPEKDKISYTSKDPANWLDPDSQITAAIGLDREDEQVKNYVEY
MVLATDSNPTTGTDITLLTLDINDHGIPEPQIIICNOSPVPQVLNITDKLSP
NNSPFOQLTHSDIYVMAEVSKEGTVALSKFKQDYDLHLSLSDHGNREQLTM
IRATVCDCHGVFNDCPRWKGFIPLCALVALLTLLALLHLVRKKRKYKPEPLL
PEATDRNVFYEGEGGEDDYDITQLHGLEARPEVVLNDVNVPIPTPYRPR
PANPDIGNFIENLKPAETDPTAPYDSLWFDYEGSGSDAASLSLITTSASDQDQD
YNYLWNGSRFKFLADNYGGGDD"

46..126

/note="signal peptide (AA -27 to -1)"

127..2511

/product="P-cadherin (AA 1 - 795)"

619..621

/note="pot. N-linked glycosylation site"

1690..1692

/note="pot. N-linked glycosylation site"

1717..1719

/note="pot. N-linked glycosylation site"

1744..1746

/note="pot. N-linked glycosylation site"

3164..3170

/note="polyA signal"

ORIGIN

Alignment Scores:

QY 319 LeuAspAlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValPro 338
 DB 976 CTTGATGCCAACCATTAACCGCTCGGAGTTTGAGCCGACAGATATGAGGCTTGGGTGCTT 1035
 QY 339 GluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsn 358
 DB 1036 GAGAACGAAGTGGCCATGAGGTACAGAGGCTGACAGTGACTGCTCGATGTCCTCCCAAC 1095
 QY 359 SerProAlaThrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThr 378
 DB 1096 TGGCCAGCGGGCTGCCACCTACCATCTGGGAGGTGATGAGGACATTTTACC 1155
 QY 379 IleThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPhe 398
 DB 1156 ATCACCACCTCACCAGACCAACCAAGCGCTCTGCAACCAAGAGGGTTGGATTTT 1215
 QY 399 GluAlaIleAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeu 418
 DB 1216 GAGGCTCAGGACCAACACACCTCTGATGAGAGTGACCAAGGCTTCCCTTTGCAAGT 1275
 QY 419 LysLeuProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaPro 438
 DB 1276 AAGCTCCGACAGCCACTGCCCGGTGGTGTCCATGTGCAAGATGTCAACGAAGCCCT 1335
 QY 439 ValPheValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluPro 458
 DB 1336 GTGTTTGTTCCTCCAGGTCTATTGAGGCCAGGAGGATCTCTATTGGGAACTG 1395
 QY 459 ValCysValThrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIle 478
 DB 1396 GTCTGCATCTATACCGCACAGGACCCAGACCAAGGAGGACCAAGAGATCAGCTACACCATC 1455
 QY 479 LeuAspProAlaGlyThrLeuAlaMetAspProAspSerGlyGlnValThrAlaVal 498
 DB 1456 TCAGAGATCCACCAACTGGCTGTGGTGGACCCAGACAGTGTTGATGATGAGTCCGCA 1515
 QY 499 GlyThrLeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetVal 518
 DB 1516 GGCATCTTGATCTGAGGACGAGCAGTGTGTGAAAACAATGCTACGAAGTCATGTT 1575
 QY 519 LeuAlaMetAspAsnGlySerProThrThrGlyThrGlyThrLeuLeuLeuThrLeu 538
 DB 1576 TTGGCCACAGACGTGGAAACCTTCCACCCAGGCACTGGACCTTCTGCTTACACTT 1635
 QY 539 IleAspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGln 558
 DB 1636 ACTGACATCAACGACCATGGCCGATCTCCGTAACCCAGGACGATCATCTGTAAACAA 1695
 QY 559 SerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerPro 578
 DB 1696 AGCCCTGTGCTCAGTGTGAACTGAACTGAAAGGACCTGTCCCAACCTCTCCCT 1755
 QY 579 PheGlnAlaGlnLeuThrAspSerAspIleTyrThrAlaGluValAsnGluGlu 598
 DB 1756 TTCAGGCGCCAGCTAACACATGACTCAGATATCTACTGGTGGCAGAGTCAAGGAGAA 1815
 QY 599 GlyAspThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHis 618
 DB 1816 GGAGACACGTCGCTTGTCCCTGAGAGTGTCTGAAACAGACAGATGATGACTTCAT 1875
 QY 619 LeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCys 638
 DB 1876 CTTTCTTTGCTCACCATGGCAACAGGAAACAGCTAACCATGATCAGGGCCACTGTGT 1935
 QY 639 AspCysHisGlyHisVal---GluThrCysProGlyProTolLysGlyGlyPheIleLeu 657
 DB 1936 GACTGCATGGCAAGTGTCAATGACTGCCACACCTCGAAGGGTGTTCATCTTC 1995
 QY 658 ProValLeuGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuVal 677
 DB 1996 CCATCTCTGGTGTCTCTGCACTGTGCTGACCTTCTACTGGCAGCTCTCTCTGTGTG 2055
 QY 678 ArgLysLysArgLysIleLysGluProLeuLeuLeuProGluAspThrArgAspAsn 697

DB 2056 AGGAAGAGAGGAAGGTCAAGAGAGCCCTTCTGCTCCAGAGATGACACGCGAGCAAT 2115
 QY 698 ValPheTyrTyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGln 717
 DB 2116 GTCTTCTATTATGGAGAGAGGGTGGTGGTGAAGAGACCCAGGACTATGATCATCCCAA 2175
 QY 718 LeuHisAspGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThr 737
 DB 2176 CTTCCACCGGGAGACTGAGAGCCAGGCTGAGTGGTCTTCCGAAACGATGTAGTGCCAC 2235
 QY 738 IleIleProThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPhe 757
 DB 2236 TTATCTCCCAACCCCATGTACGAGCCCGCCGACCAACCCAGATGAATCGGAACTTC 2295
 QY 758 IleIleGluAsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeu 777
 DB 2296 ATCATCGAGAACCTGAGACCTGCCAACAACCTACTGCCCGCCCTACGACTCCCTG 2355
 QY 778 LeuValPheAspTyrGluGlySerGlySerAspAlaIleSerLeuSerSerLeuThrSer 797
 DB 2356 ATCGTGTGTTGACTACGAGGCGAGCGGCTCTGATCGCCCTCCCTGAGCTCCCTCAC 2415
 QY 798 SerAlaSerAspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLys 817
 DB 2416 TCCGCTCTCGACCAAGATCAGACTACACTACTTAAACGATGGGAAAGTCGATTCCAG 2475
 QY 818 LysLeuAlaAspMetTyrGlyGlyGlyAsp 829
 DB 2476 AAACTGGCGCATGTATGTTGGTGGGCTGAGGATGAC 2511
 RESULT 8
 BC052189
 LOCUS
 DEFINITION Mus musculus cadherin 3, mRNA (CDNA clone IMAGE:6394535), partial cds.
 ACCESSION BC052189
 VERSION BC052189.1 GI:30704940
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 1 (bases 1 to 2993)
 STRAUSBERG, R.L., FEINGOLD, E.A., GROUSE, L.H., DERGE, J.G., KLAUSNER, R.D., COLLINS, F.S., WAGNER, L., SHENMEN, C.M., SCHULER, G.D., ALTSCHUL, S.F., ZEEBERG, B., BUETOW, K.H., SCHAEFER, C.F., BHAT, N.K., HOPKINS, R.F., JORDAN, H., MOORE, T., MAX, S.I., WANG, J., HSIEH, F., DIATCHENKO, L., MARUSINA, K., FARMER, A.A., RUBIN, G.M., HONG, L., STAPLETON, M., SOARES, M.B., BONAIDO, M.F., CASAVANT, T.L., SCHEETS, T.E., BROWNSTEIN, M.J., UEDIN, T.B., TOSHIYUKI, S., CARNINCI, P., PRANGE, C., RAHA, S.S., LOQUELLANO, N.A., PETERS, G.J., ABRAMSON, R.D., MULLAHL, S.J., BOSAK, S.A., MCBWAN, P.J., MCKERNAN, K.J., MALEK, J.A., GUNARATNE, P.H., RICHARDS, S., WORLEY, K.C., HALE, S., GARCIA, A.M., GAY, L.J., HULYK, S.W., VILLALON, D.K., MUZYNY, D.M., SODERGREN, E.J., LU, X., GIBBS, R.A., FAHEY, J., HELTON, E., KETTEMAN, M., MADAN, A., RODRIGUES, S., SANCHEZ, A., WHITING, M., MADAN, A., YOUNG, A.C., SHEVCHENKO, Y., BOUFFARD, G.G., BIRKESLEY, R.W., TOUCHMAN, J.W., GREEN, E.D., DICKSON, M.C., RODRIGUEZ, A.C., GRIMWOOD, J., SCHMUTZ, J., MYERS, R.M., BUTTERFIELD, Y.S., KRZYWINSKI, M.I., SKALSKA, U., SMALIUS, D.E., SCHNERCH, A., SCHEIN, J.E., JONES, S.J. and MARRA, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 JOURNAL 2238057
 MEDLINE 12477932
 REFERENCE
 2 (bases 1 to 2993)
 STRAUSBERG, R.
 Direct Submission
 Submitted (01-MAY-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

REMARK
COMMENT

USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman, Ph.D.
CDNA Library Preparation: ResGen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalilus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAP Plate: 109 Row: 9 Column: 19
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES

source

Location/Qualifiers
1..2993
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6394535"
/tissue_type="Embryo, day 9 mouse (C57BL/6 background) otocysts"
/clone_lib="NTH_MGC_130"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6.1"
<1..2993
/gene="Cdh3"
/note="synonym: P-cadherin"
/db_xref="LocusID:12560"
/db_xref="MGI:88356"
<1..2264
/gene="Cdh3"
/codon_start=3
/product="Cdh3 protein"
/protein_id="AAH52189.1"
/db_xref="GI:30704941"
/db_xref="LocusID:12560"
/db_xref="MGI:88356"
/translation="DIIMLRGTVQGGKDAHSPPTILRRKRKRWMPPIFVPENGK
GPPFQRNLQSKNRGKTKIFYITGPGADSPPEGVFTIEKSGWLLHMPDLREKIV
KYLEYHVAENGASVEEPMNIIITVDQNDKPKFTQDTFRGSVLEGVMPGTSVMQV
TATDEDDAVNTYNGWVAYSITHSQEPKPHDLMTIKSTGTISVSSGLDRKVPYR
LTQVATDMCEGSTTTAEAVVQILDANDNAPEPEPOKYEAWPENEVGVORLTVTD
LDVNSPAWKTAVHVGDDGDHFTITHTPBTNQGLVITKGLDPAQOHLIVSVT
NEAPPVAKLTATAVVHVVDNEAPVFPVPSKVEIAQEGISIGELVCITYAQDDPK
EDKQSTILIRDPANLVHVPDPTGAIGLREDEQFVKNVYEVWVLATDSGNPP
TTGTLLLTLDINHGPIPEPQIIICNQSPVQLNITDKDLSFNPSFPQALVTH
DSDIYMAVSEKGTVALSKFLQDTYDLHLSLDHGRNQLMTIRATCDCHGQ
VFNDGPRPMGGGFIPLIGAVALLTLLALLLVKKRKKVEPIALLPDDTDNDVFI
YGEEGGEEDQDITQLHGLEARPEVLRNDVPTFIPTMYRPRRANDEIGNFI
IENLKAANTDPTAPPYDSLLVFDYEGSGDAASLSLTTISASDQDDQYINLEWGRF
KKLADYVGGEDD"

gene

CDS

Alignment Scores:
Pred. No.: 3,77e-252 Length: 2993
Score: 3433.50 Matches: 639
Percent Similarity: 94.01% Conservative: 51
Best Local Similarity: 87.06% Mismatches: 43
Query Match: 78.59% Indels: 1
DB: 10 Gaps: 1
US-09-916-849A-1 (1-829) x BC052189 (1-2993)
QY 97 ProSerLysArgIleLeuArgArgHisLysArgAspTTPValValAlaProIleSerVal 116
Db 60 CCACCCACCCGCATCTTAAGGAGACGAAGAGAGAGTGGTGATGCCACCATATTCGTC 119
QY 117 ProGluAsnGlyLysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAsp 136
Db 120 CCCGAGATGGCAAGGGTCCCTTCCTCAGAGGCTGAATCAGCTCAAAATCTAATAAGGAC 179
QY 137 ArgAspThrLysIlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGly 156
Db 180 AGAGGCACCAAGATTTTCTACAGCATCAGAGGCTGGGCGAGACAGTCCCCCGAAGA 239
QY 157 ValPheAlaValGluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGlu 176
Db 240 GTCTTCACCATAGACAGAGTGGCGTGGCTGTGTGTTCATATGCCATCGACAGGAG 299
QY 177 GluIleAlaLysTyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGlu 196
Db 300 AGATTGTCACTACAGCTTATGCGCCAGCTTATGCGCCAGCTTATGCGCCAGCTTCTGTAG 359
QY 197 AspProMetAsnIleSerIleValThrAspGlnAsnAspHisLysProLysPheThr 216
Db 360 GAGCCCATGACATATCCATTCATGTGACAGACCAAGATGACACAGCCCAAGTTCAT 419
QY 217 GlnAspThrPheArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGln 236
Db 420 CAAGACACCTTCAGAGGAGTGTTCCTGGAGGAGTAATGCGCTGGCAGTCTCTGTAG 479
QY 237 ValThrAlaThrAspGluAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSer 256

misc_feature

105..389
/gene="Cdh3"
/note="cadherin; Region: Cadherin domain"

misc_feature

480..749
/gene="Cdh3"
/note="CA; Region: Cadherin repeats. Cadherins are glycoproteins involved in Ca²⁺-mediated cell-cell adhesion. Cadherin domains occur as repeats in the

extracellular regions which are thought to mediate cell-cell contact when bound to calcium"
/db_xref="CDD:smart00112"
768..1064
/gene="Cdh3"
/note="cadherin; Region: Cadherin domain"
/db_xref="CDD:pfam00028"
1158..1409
/gene="Cdh3"
/note="CA; Region: Cadherin repeats. Cadherins are glycoproteins involved in Ca²⁺-mediated cell-cell adhesion. Cadherin domains occur as repeats in the extracellular regions which are thought to mediate cell-cell contact when bound to calcium"
/db_xref="CDD:smart00112"
1422..1685
/gene="Cdh3"
/note="cadherin; Region: Cadherin domain"
/db_xref="CDD:pfam00028"

misc_feature

1845..2252
/gene="Cdh3"
/note="Cadherin C term; Region: Cadherin cytoplasmic region. Cadherins are vital in cell-cell adhesion during tissue differentiation. Cadherins are linked to the cytoskeleton by catenins. Catenins bind to the cytoplasmic tail of the cadherin. Cadherins cluster to form foci of homophilic binding units. A key determinant to the strength of the binding that it is mediated by cadherins is the juxtamembrane region of the cadherin. This region induces clustering and also binds to the protein p120ctn"
/db_xref="CDD:pfam01049"

ORIGIN

Alignment Scores:
Pred. No.: 3,77e-252 Length: 2993
Score: 3433.50 Matches: 639
Percent Similarity: 94.01% Conservative: 51
Best Local Similarity: 87.06% Mismatches: 43
Query Match: 78.59% Indels: 1
DB: 10 Gaps: 1
US-09-916-849A-1 (1-829) x BC052189 (1-2993)
QY 97 ProSerLysArgIleLeuArgArgHisLysArgAspTTPValValAlaProIleSerVal 116
Db 60 CCACCCACCCGCATCTTAAGGAGACGAAGAGAGAGTGGTGATGCCACCATATTCGTC 119
QY 117 ProGluAsnGlyLysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAsp 136
Db 120 CCCGAGATGGCAAGGGTCCCTTCCTCAGAGGCTGAATCAGCTCAAAATCTAATAAGGAC 179
QY 137 ArgAspThrLysIlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGly 156
Db 180 AGAGGCACCAAGATTTTCTACAGCATCAGAGGCTGGGCGAGACAGTCCCCCGAAGA 239
QY 157 ValPheAlaValGluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGlu 176
Db 240 GTCTTCACCATAGACAGAGTGGCGTGGCTGTGTGTTCATATGCCATCGACAGGAG 299
QY 177 GluIleAlaLysTyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGlu 196
Db 300 AGATTGTCACTACAGCTTATGCGCCAGCTTATGCGCCAGCTTATGCGCCAGCTTCTGTAG 359
QY 197 AspProMetAsnIleSerIleValThrAspGlnAsnAspHisLysProLysPheThr 216
Db 360 GAGCCCATGACATATCCATTCATGTGACAGACCAAGATGACACAGCCCAAGTTCAT 419
QY 217 GlnAspThrPheArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGln 236
Db 420 CAAGACACCTTCAGAGGAGTGTTCCTGGAGGAGTAATGCGCTGGCAGTCTCTGTAG 479
QY 237 ValThrAlaThrAspGluAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSer 256

Db 480 GTGACAGCCACAGATGAGGACGATGCTGTCAACACCTTACAATGGGCTGGTCTTACTCC 539
 QY 257 ILeHisSerGlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThr 276
 Db 540 ATCCATAGCCCAAGAGCGAAGGACGACACGACCTCATGTTCCATCATCAATAAAGCAGC 599
 QY 277 GlyThrIleSerValIleSerSerGlyLeuAspArgGluLeuValProGluTyrThrLeu 296
 Db 600 GGAACCATTAGCGTCATATCCAGTGGCCCTGGACGAGAGAAAGTCCCTGAGTACAGACTG 659
 QY 297 ThrIleGlnAlaThrAspMetAspGlyAspGlySerThrThrAlaValAlaVal 316
 Db 660 ACCGTCAGGCCACAGACATGAGTGGAGAGGCTCTTACCACGACGACGAGCGCTGGT 719
 QY 317 GluIleLeuAspAlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHis 336
 Db 720 CAAATCCTTGATGCCAACGATAACGCTCCGGAGTTTGAGCCGACAGATGATGAGGCTGG 779
 QY 337 ValProGluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAla 356
 Db 780 GTCCCTGAGAACCAAGTGGCCATGAGGTACAGAGGCTGACAGTCACTGATCTCGATGTC 839
 QY 357 ProAsnSerProAlaThrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHis 376
 Db 840 CCCAACTCGCAGCGTGGCGTCCACTACCATCGTGGGAGGTGATGATGGGACCAT 899
 QY 377 PheThrIleThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeu 396
 Db 900 TTCACCATCACCACTCACCAGAGACCAACCAAGCGCTCCTGACAAACCAAGAGGGTTG 959
 QY 397 AspPheGluAlaLysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPhe 416
 Db 960 GATTTTGAGGCTCAGGACCAACACACCCCTGTATGAGAAGTGCACCAACGAGGCTCCCTTT 1019
 QY 417 ValLeuLysLeuProThrSerThrAlaThrIleValValHisValGluAspValAsnGlu 436
 Db 1020 GCGATGAAGTCCCGACACCACTGCCACCGTGGTGGTCCATGTAAGATGTCAACGAA 1079
 QY 437 AlaProValPheValProSerLysValValGluValGlnGluGlyIleProThrGly 456
 Db 1080 GCCCTGTGTTGTTCCACTTCCAAAGTCACTTAGGCCCCAGGAGGCATCTCTATTGGG 1139
 QY 457 GluProValCysValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyr 476
 Db 1140 GAATCGTGTGATCATATCCACAGACAGGACAGGACAGGAGGACCAAGATCAGGTAC 1199
 QY 477 ArgIleLeuArgAspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThr 496
 Db 1200 ACCATCTCGAGAGATCCAGCCAACTGGCTTGTGGAGCCACAGACAGTGGTCAATAACT 1259
 QY 497 AlaValGlyThrLeuAspArgGluAspGluGlnPheValArgAsnIleTyrGluVal 516
 Db 1260 GCGCAGGAGCATCTGGATCGTGAGGACGAGCAGTGTGTGAATAACAATGTCTAGAAATC 1319
 QY 517 MetValLeuAlaMetAspAsnGlySerProThrThrGlyThrGlyThrLeuLeuLeu 536
 Db 1320 ATGGTTTTGGCCACAGACAGATGGAACCTCCACCCAGCGCACTGGGACCTCCTGCTT 1379
 QY 537 ThrLeuIleAspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCys 556
 Db 1380 ACATTTACTGATCAACAGCAGCATGGCCGATPCCCTGAACCCAGGACAGATCATCATCTGT 1439
 QY 557 AsnGlnSerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThr 576
 Db 1440 AACCAAGCCCTGTGCTCAAGTGTGAACATCACTGACAGGACCTGTCCCCCACTCC 1499
 QY 577 SerProPheGlnAlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsn 596
 Db 1500 TCCCTTTTCCAGCCCGCAGTAAACATCACTCAGATATCTACTGGATGGGAGAGTCAAGC 1559
 QY 597 GluGluGlyAspThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAsp 616
 Db 1560 GAGAAAGGAGACACCGTGGCCTTGTCCCTGAAGAAGTTCCTGAAACAGACACAGTATGAC 1619

QY 617 ValHisLeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThr 636
 Db 1620 TTGCATCTTTCTTTGTCTGACCATGGCAACAGGAACAGCTAACCATGATCAGGGCCACT 1679
 QY 637 ValCysAspCysHisGlyHisVal---GluThrCysProGlyProTyrPlysGlyGlyPhe 655
 Db 1680 GTGTGTGACTGTCATGGCCAAAGTGTCAATGACTGCCACAGACCTCGAAGGGTGGTTTC 1739
 QY 656 IleLeuProValLeuGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeu 675
 Db 1740 ATCTTCCCATCTCTGGTGTCTCTGCACTGTGACCTTCTACTGGCACTCTCTCTG 1799
 QY 676 LeuValArgLysLysArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArg 695
 Db 1800 TTGGTGAGAGAGAGAGAGAGTCAAGAGAGCCCTTCTGTCTCCAGAGATGACACGCGA 1859
 QY 696 AspAsnValPheTyrTyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIle 715
 Db 1860 GACATGCTTCTATATGGAGAGAGGGTGTGTGAAGAGGACAGACTATGATC 1919
 QY 716 ThrGlnLeuHisArgGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAla 735
 Db 1920 ACCCAACTCCACCGGGAGTGGAGGCCAGGCTGAGGTGGTCTCCGAAAACGATGTAGTG 1979
 QY 736 ProThrIleIleProThrProMetTyrArgProArgProAlaAsnProAspGluIleGly 755
 Db 1980 CCAACTCTCATCCCAACCCCTGTTACCGACCCCGCCGCCAACCCAGATGAATCGGG 2039
 QY 756 AsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAspProThrAlaProTyrAsp 775
 Db 2040 AACTTTCATCATCGAGAACTGAAGGCTGCCAACACTGACCTACTGCCCGCCCTACGAC 2099
 QY 776 ThrLeuLeuValPheAspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeu 795
 Db 2100 TCCCTGCTGGTGTGTGATGACGAGGAGCGCTCTGTATGCGGCTCCTCTGAGCTCCCTC 2159
 QY 796 ThrSerAlaSerAspAspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArg 815
 Db 2160 ACCACTCGGCTCCGACGAGTACAGTACAACTTAACTTAACTTAACTTAACTTAACTTAACT 2219
 QY 816 PheLysLysLeuAlaAspMetTyrGlyGlyGlyGluAspAsp 829
 Db 2220 TTCAAGAACTGGCGGACATGTATGGTGGCGGTGAGGATGAC 2261
 RESULT 9
 LOCUS GGCADHB 2436 bp mRNA linear VRT 20-JUL-1995
 DEFINITION Chicken mRNA for B-cadherin.
 ACCESSION X58518
 VERSION XS8518.1 GI:63113
 KEYWORDS B-cadherin; cadherin; transmembrane protein.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 2436)
 AUTHORS Napolitano, E.W., Venstrom, K., Wheeler, E.F. and Reichardt, L.F.
 TITLE Molecular cloning and characterization of B-cadherin, a novel chick
 cadherin
 JOURNAL J. Cell Biol. 113 (4), 893-905 (1991)
 MEDLINE 91225083
 PUBMED 2026653
 REFERENCE 2 (bases 1 to 2436)
 AUTHORS Napolitano, E.W.
 TITLE Direct Submission
 JOURNAL Submitted (01-MAR-1991) E.W. Napolitano, Howard Hughes Medical
 Institute, University of California, 3rd and Parnassus Sts, San
 Francisco CA 94143-0724, USA
 FEATURES
 source 1. 2436
 /organism="Gallus gallus"

```

/mol_type="mRNA"
/db_xref="taxon:9031"
/clone="CAD-51"
/tissue_type="brain"
/clone_lib="lambda gt11 E13 chick brain cdna"
/dev_stages="embryonic day 13"
1..2436
/note="ORF1"
B-cadherin
/evidence=experimental
<1..2181
/note="ORF1"
/codon_start=1
/evidence=experimental
/product="B-cadherin"
/protein_id="CAA1408.1"
/db_xref="GI:63114"
/db_xref="GOA:P33145"
/db_xref="SWISS-PROT:P33145"
/translacion="DWIPPIKVPENERGPPKXNLVOIKSNRDEAKIFYISITQOGAD
APPGITIEKETGMKVTOPLDREHINKVHLXSHVSENGKPEVERPMEIIVTVTON
DNKQFTQEVPRGSVPKALPGTSVMVNATDADDDVETVNGVYAVSILSOEEREHP
HMPYVNAATGTSVIAGLDRERVRETLTQWADLDGQGLTTALAVISITVDNDA
PEFPIKTEAAVPENEAEVARLATDDEPHTPAWRAVISVRNGEGGAFTITDP
ASNEGLVRLTAKGLDEYKARQFVLHVAVNEAPFAIKLPTATATVMSVEDVNEAPFD
PPLRAQVEDPVLQPLASYAQDPDRAQQQRIKYVMGSDPAGWLAHPENGLIAR
BQLDRESPTKNSYVAVLAVDDGLPPTGTTGLTLLDNDVHDGPEPEPRIVICN
RSPYVOLTITDRLPNTGTFRAELSHGSDSWAVEVNGDGTVALWLEPELQNLV
SVLRLPDROGKQDVTIRAQVCDQGRVSCAKPRVDTVLAVGLAVLALLV
LILLLLLVRRKKVKEILLPEDTRONIFYEBEGGEBEDQDIDLSQLHKGDAPE
VLRNDAPPLMAAPQYPRFANPDEIGNFIDENIKAADTPTAPPYDLSLVFVYEGG
SEATSLSSLNSSADSDQDYDLYNEWGNRPFKLAELYGGGEDEE"
polyA_signal
polyA_site
2397..2402
2416
ORIGIN
Alignment Scores:
Pred. No.: 7.74e-185 Length: 2436
Score: 2549.50 Matches: 480
Percent Similarity: 80.19% Conservative: 103
Best Local Similarity: 66.02% Mismatches: 137
Query Match: 58.35% Indels: 7
DB: 5 Gaps: 5
US-09-916-849A-1 (1-829) x GGCADHB (1-2436)
Qy 108 AspTrrpValValAlaProIleSerValProGluAsnGlyLysGlyProPheProGlnArg 127
Db 1 GATTGGGTGATCCCTCCCAATTAAAGTTCCTCCGAGATGAGAGGGTCCCTTCCCCAAAAAC 60
Qy 128 LeuAsnGlnLeuLysSerAsnLysAspArgAspThrLysilePheTyrSerIleThrGly 147
Db 61 TTGTTTCAGATCAAAATCCACCGGACCGGAGCAAGATTTTCTACAGCATCACCGGG 120
Qy 148 ProGlyAlaAspSerProProGluGlyValPheAlaValGluLysGluThrGlyTrrpLeu 167
Db 121 CAGGGGGCGGATGCTCCCCCAGGGCATCTTACCATTGAGAGGAGACAGGCTGGATG 180
Qy 168 LeuLeuAsnLysProLeuAspArgGluGluIleAlaLysTyrGluLeuPheGlyHisAla 187
Db 181 AAGGTGACGACGCGCTGGACCGGAGCACATCAACAAGTACACACCTCTACTCCCATGCC 240
Qy 188 ValSerGluAsnGlyAlaSerValGluAspProMetAsnIleSerIleValThrAsp 207
Db 241 GTGTCGAGAGATGCAACCCGTCGAGGACCGGATGAGATCATAGTGACGGTGACGGAC 300
Qy 208 GlnAsnAspHisLysProLysPheThrGlnAspThrPheArgGlySerValLeuGluGly 227
Db 301 CAAAATGACACAAAGCCCACTTACACAGAGAGGTGTTCCGGGGCTCCGTGCCTGAAGGC 360
Qy 228 ValLeuProGlyThrSerValMetGlnValThrAlaThrAspGluAspAlaIleTyr 247
Db 361 GCGTGCAGGCACTTCGGTGATCGGGTGAAACGCCACCGCATGATGATGTGGAG 420
248 ThrTyrAsnGlyValValAlaTyrSerIleHisSerGlnGluProLysAspProHisAsp 267
421 ACCTACACAGGTGTCTATCCTCCTACTCCTCCTCAGCAGGAGCGCGGAGCCCATCC 480
268 LeuMetPheThrIleHisArgSerThrGlyThrIleSerValIleSerGlyLeuAsp 287
481 CACATGTTTACCGTCAACCGCGCCACCGCACCTCAGCGTCTATCGCAGCGGCTCGAC 540
288 ArgGluValValProGluTyrThrLeuThrIleGlnAlaThrAspMetAspGlyAspGly 307
541 CGGGAGCGGTGCGGGAGTACACGCTGACATGACGCGGCTGACCTGACGCGGAGGGG 600
308 SerThrThrAlaValAlaValGluIleLeuAspAlaAsnAspAlaProMet 327
601 CTGACCACACAGCGCTGCTGTGATTGAGATCAAGACGCTCAATGACACGCCCTGAG 660
328 PheAspProGlnLysTyrGluAlaHisValProGluAsnAlaValGlyHisGluValGln 347
661 TTGACCCCAAAACGATGATGAGCGGTGCTGCGCAGAGATGAGCGCGAGGTGGCC 720
348 ArgLeuThrValThrAspLeuAspAlaProAsnSerProAlaTrrpAlaThrTyrLeu 367
721 CGGTTGCGCACCGACGCTGATGAGCGGCACACACACCGCGTGGCGTCCCTCTACTCC 780
368 IleMetGlyGlyAspAspGlyAspHisPheThrIleThrThrHisProGluSerAsnGln 387
781 ATCGTTCGTGGCAATGAGGCGGTGCTTACCATCATCACCCTGACCTCCAGCAACGAG 840
388 GlyIleLeuThrThrArgLysGlyLeuAspPheGluAlaLysAsnGlnHisThrLeuTyr 407
841 GCGCTCTCGCACCGCCCAAGGCTGTGGACTACGAGGCCAAGCGGAGTTCGTGTCTCCAC 900
408 ValGluValThrAsnGluAlaProPheValLeuLysLeuProThrSerThrAlaThrIle 427
901 GTGGCGGTGCTCAATGAAGTCCCTTCGCCATCAAGTCCACACGCCACGCCACCGGTG 960
428 ValValHisValGluAspValAsnGluAlaProValPheValProSerLysValVal 447
961 ATGCTCAGTGTGGAGATGTAATGAGCGCGCGCTTTCAGACCGCGCTGGGTGGCC 1020
448 GluValGlnGluLysLeuProThrGlyGluProValCysValTyrThrAlaGluAspPro 467
1021 CAGGTCCCGGAGACGTGCGCTGGCGGACGCCCTCTCTCTATACACACGAGCCCT 1080
468 AspLys--GluAsnGlnLysIleSerTyrArgIleLeuArgAspProAlaGlyTrrpLeu 486
1081 GACAGGCGCCACGACGAGCGCATCAAGTACGTGATGGGCGAGTACCCCGCGGCTGGCTG 1140
487 AlaMetAspProAspSerGlyGlnValThrAlaValGlyThrLeuAspArgGluAspGlu 506
1141 GCGGTGCACCCCGAGAACCGCTCATCAGCAGCGGAGCAGCTGACCGTGGAG--TCC 1197
507 GlnPheValArgAsnAsnIleTyrGluValMetValLeuAlaMetAspAsnGlySerPro 526
1198 CGTTTACCAAGAACAGCACCTACGTGGCGGTGCTGTGGCTGTGGATGACGATGCCA 1257
527 ProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAspValAsnAspHisGlyPro 546
1258 CCGCGCACAGGACCGGACGCTGCTCTCCTCCTCCTGCTGATGTCAATGACACCGCCCT 1317
547 ValProGluProArgGlnIleThrIleCysAsnGlnSerProValArgHisValLeuAsn 566
1318 GAGCGGAGCGCCCGGACATCTGCTGTATACCGCCCTGTGCCCCAGGTGCTGACC 1377
567 IleThrAspLysAspLeuSerProHisThrSerProPheGlnAlaGlnLeuThrAsp 586
1378 ATCAGCGACCGGACCTGCCCCCAACACCGGGCGCTTCCGTGCGGAGTGGACCGGC 1437
587 SerAspIleTyrTrrpThrAlaGluAsnGluGlyAspThrValValLeuSerLeu 606
1438 TCTGGGAGCAGCTGGGCGGTGGAGTTGCCAATGAGGAGTACACCTTGTGCTGTGGCTG 1497

```


QY 281 ValleserSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
 Db 1076 GTGTTCACCACTGGCGTGGACGAGAGATTTCCTTACCTATACCTGGTGGTTCAGCT 1135
 QY 301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValGluLeuAsp 320
 Db 1136 GCTGACCTTCAAGTGGGGTAAAGCACACACACAGCTGTGATCACAGTCACTGAC 1195
 QY 321 AlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
 Db 1196 ACCAACGATAATCTCCGATCTTCAATCCACACGACGACGAGGTCAGGTGCCTGAGAAC 1255
 QY 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
 Db 1256 GAGCTAACGTCGTAATCACACACTGAAGTGAAGTGTGATGATGATGATGATGATGAT 1315
 QY 361 AlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
 Db 1316 GCGTGGGAGGCTGTATACACCATATG--AATGATGATGATGATGATGATGATGATGAT 1372
 QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
 Db 1373 ACAATCCAGTGAACACGATGGCATTTTGAACACACGAAAGGCTTGGATTTTGAAGCC 1432
 QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
 Db 1433 AAGCAGCAGTACATCTACAGCTAGCAGTGAAGTGTGATGATGATGATGATGATGATGAT 1492
 QY 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
 Db 1493 ACCACCTCCACAGCACCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1552
 QY 441 ValProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
 Db 1553 GTGCTCTGTAAGAGAGTGAAGTGTCCGAGACCTTGGCGTGGCCACGAAATACACA 1612
 QY 461 ValTyrThrAlaGluAspProAspLys--GluAsnGlnLysIleSerTyrArgIleLeu 479
 Db 1613 TCCTACACTGCCAGGAGGAGATTTGACACAGTGAAGACACAGCAGTACACAGCCCTAATCAT 1792
 QY 480 ArgAspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGly 499
 Db 1673 AGAGACTGCCAAGTGGTGGAGATTAATCCGACACTGTCGATGATGATGATGATGATGATGAT 1732
 QY 500 ThrLeuAspArgGluAspGluGlnPheValArgAsnIleTyrGluValMetValLeu 519
 Db 1733 GAGCTGGACAGGAGGATTTGACACAGTGAAGACACAGCAGTACACAGCCCTAATCAT 1792
 QY 520 AlaMetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeu 539
 Db 1793 GCTACAGACATGTTCTCCAGTTGCTACTGGAACAGGACACTTCTGCTGATCCTGCT 1852
 QY 540 AspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSer 559
 Db 1853 GATGTGATGACAGCGCCCATACAGACCTCGAATATATCTTCTGTGAGAGGAT 1912
 QY 560 ProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPhe 579
 Db 1913 CCAAAGCTCAGGTCATAAATCATCATGATGATGATGATGATGATGATGATGATGATGATGAT 1972
 QY 580 GlnAlaGlnLeuThrAspAspSerAspIleTyrTyrThrAlaGluValAsnGlu--Glu 598
 Db 1973 ACAGCAGAACCTAACACAGCGGGCGAGTGCACACTGGACCATTCAGTACACAGCCCAAC 2032
 QY 599 GlyAspThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHis 618
 Db 2033 CAAGAATCTATATTTGAAGCCAAAGATGGCTTAGAGTGGGTGATACAAATCAAT 2092
 QY 619 LeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCys 638
 Db 2093 CTAAGTCTAGGATAACCAAGATGAAGACCAAGTGCACCTAGAGGTTCAGCGTGTGT 2152

QY 639 AspCysHisGlyHisValGluThrCys-----ProGlyProTyrPheGlyGlyPhe--- 655
 Db 2153 GACTGTGTAAGGCGCGCGCGCTGTGTAGAAAGGCACAGCCTGTCAAACAGGATTGCAA 2212
 QY 656 -----IleLeuProValLeuGlyAlaValLeuAlaLeuLeuPheLeuLeuVal 672
 Db 2213 ATTCTGCGCATTTCTGGGATTTCTGGAGAAATCTTCTTCTTCTTCTTCTTCTTCTTCT 2272
 QY 673 LeuLeuLeuLeuValArgLysLysArgLysIleLysGluProLeuLeuLeuProGluAsp 692
 Db 2273 CTCTGCTCTTCTTCTGGAGAGAGCGGTGTCAAAGAGCCCTTACTGCCCCAGAGGAT 2332
 QY 693 AspThrArgAspAsnValPheTyrGlyGluGluGlyGlyGlyGlyGluAspGlnAsp 712
 Db 2333 GACACCGGACACACTTTATTACTATGATGAAGAGGAGGCGGAGAGAGACAGGAC 2392
 QY 713 TyrAspIleThrGlnLeuHisArgGlyLeuGluAlaArgProGluValValLeuArgAsn 732
 Db 2393 TTTGACTTGACCGACGTGCACAGGGCTTGACGCTCGSCCTGAAGTGACT--CGTAAC 2449
 QY 733 AspValAlaProThrIleIleProThrProMetTyrArgProArgProAlaAsnProAsp 752
 Db 2450 GAGTTTGACCAACCTCATGATGTCCCGCTATCTTCCCGCCTGCCAATCCCGAT 2509
 QY 753 GluIleGlyAsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAspProThrAlaPro 772
 Db 2510 GAAATTGGAATTTTATTGATGAATCTGAAGCGGCTGATCTGACCCACAGCCCG 2569
 QY 773 ProTyrAspThrLeuLeuValPheAspTyrGluGlySerGlySerAspAlaSerLeu 792
 Db 2570 CTTTATGATTTCTGCTGCTGTTGACTATGAAGAGCGGTTCGGAAGCTCTAGTCTG 2629
 QY 793 SerSerLeuThrSerSerAlaSerAspGlnAspGlnAspTyrAspTyrLeuAsnGluTyr 812
 Db 2630 AGCTCCTGAACTCTCAGAGTCAGACAAAGACAGGACTATGACTTGTGAACGATGG 2689
 QY 813 GlySerArgPheLysLysLeuAlaAspMetTyrGlyGlyGlyGluAspAsp 829
 Db 2690 GCGAATCGTTCAAGAGCTGCTGACATGTACGAGCGCGGCGAGAGCAG 2740

RESULT 11
 HSCAD HSCAD 4778 bp mRNA linear PRI 01-JUL-1999
 LOCUS H.sapiens mRNA for E-cadherin.
 DEFINITION Z13009
 ACCESSION Z13009
 VERSION 1 GI:31072
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 REFERENCE 1 (bases 1 to 4778)
 AUTHORS Bussemakers,M.J., van Bokhoven,A., Mees,S.G., Kemler,R. and Schalken,J.A.
 TITLE Molecular cloning and characterization of the human E-cadherin cDNA
 JOURNAL Mol. Biol. Rep. 17 (2), 123-128 (1993)
 MEDLINE 93211394
 PUBMED 8459805
 REFERENCE 2 (bases 1 to 4778)
 AUTHORS Bussemakers,M.J., Giroidi,L.A., van Bokhoven,A. and Schalken,J.A.
 TITLE Transcriptional regulation of the human E-cadherin gene in human prostate cancer cell lines: characterization of the human E-cadherin gene promoter
 JOURNAL Biochem. Biophys. Res. Commun. 203 (2), 1284-1290 (1994)
 MEDLINE 94380041
 PUBMED 8093045
 REFERENCE 3 (bases 1 to 4778)
 AUTHORS Bussemakers,M.
 TITLE Direct Submission
 JOURNAL Submitted (22-JUN-1992) Bussemakers M., University Hospital Nijmegen, Geert Grooteplein 16, Nijmegen, The Netherlands
 FEATURES
 source 1..4778


```

1339 AAGCAGCAGTACATCTACACGTAGCAGTACGAATGTGGTACCTTTGAGGTCTCTCTC 1398
421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
1399 ACCACCTCCACAGCAGCCGTCACCGTGGATGTGTGATGTAAGAACGCCCATCTTT 1458
441 ValProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
1459 GTGCCTCTCTGAAGAAGAGAGTGAAGTGTCCGAGGACITTCGCGGCGCCAGGAATCACA 1518
461 ValTyrThrAlaGluAspProAspLys---GluAsnGlnIleSerTyrArgIleLeu 479
1519 TCCTACACTGCCAGCCAGCAGACACATTTATGGAACAGAAATAACATATCCGATTTGG 1578
480 ArgAspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGly 499
1579 AGAGACACTGCCAAGTCTGAGATTAATCCGACACTGTGCGCATTTTCACTCGGGCT 1638
500 ThrLeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeu 519
1639 GAGCTGGACAGGAGGATTTTGACGACGTGAAGAACACAGCCTACACAGCCCTAATCATA 1698
520 AlaMetAspAsnGlySerProProThrThrClyThrGlyThrLeuThrLeuLeu 539
1699 GCTACAGCAATGGTCTCCAGTTGCTACTGGAACAGGACACTCTGCTGATCTCTCT 1758
540 AspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSer 559
1759 GATGTGATGATCAACAGCCGCCCATACACAGACCTCGAATATATCTCTGTGAGAGGAAT 1818
560 ProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPhe 579
1819 CCAAGCCTCAGGTCATAAATCATTTGATGACAGACCTTCTCCCAATACATCTCCCTC 1878
580 GlnAlaGlnLeuThrAspSerAspIleTyrThrAlaGluValAsnGlu---Glu 598
1879 ACAGCAGAACTAACACAGCGGCGAGTCCCACTGGACCATTCAGTACACAGCCCAACC 1938
599 GlyAspThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHis 618
1939 CAAGAACTATCATTTTGAACCCAAAGATGCGCTTAGAGTGGGTGACTACAAATCAAT 1998
619 LeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCys 638
1999 CTCAGCTCATGATACCAATTAAGACCAAGTGTACCACTTAGACCTTAGAGTACGCGTGT 2058
639 AspCysHisGlyHisValGluThrCys-----ProGlyProTyrLysGlyGlyPhe--- 655
2059 GACTGTGAAGGGCGCGCGCGCTCTGTAGGAAGGCACAGCGCTGTGAAGCAGGATTGCAA 2118
656 -----IleLeuProValIleGlyAlaValLeuAlaLeuLeuPheLeuVal 672
2119 ATTCTGCGCATCTCTGGGATCTTGGAGAAATCTTCTGTTGCTTAATCTGATCTCTG 2178
673 LeuLeuLeuValValArgLysArgLysIleLysGluProLeuLeuLeuProGluAsp 692
2179 CTCCTGCTGTTCTTCGGAGGAGCGGTGTCAAGAGCCCTTACTGCCCCCAGAGGAT 2238
693 AspThrArgAspAsnValPheTyrTyrGlyGluGlyGlyGlyGluAspGlnAsp 712
2239 GACACCCGGGCAACGTTTATTACTATGATGAAGAGGCGGAGAGGAGGACCGAGC 2298
713 TyrAspIleThrGlnLeuHisArgGlyLeuGluAlaArgProGluValValLeuArgAsn 732
2299 TTTGACTTGAGCCAGCTGCACAGGGGCTGACGCTGCGCTGAAGTACT---CGTAAC 2355
733 AspValAlaProThrIleProThrProMetTyrArgProArgProAlaAsnProAsp 752
2356 GACGTTGACCAACCTCATGAGTGTCCCGCGGTATCTTCCCGCCCTGCCAATCCCGAT 2415
753 GluIleGlyAsnPheIleGluAsnLeuLysAlaAlaAsnThrAspProThrAlaPro 772

```

```

2416 GAATTGGAAATTTTATTGATGAAATCTGAAAGCGCGGTGATCTGATGACCCACAGCCCGC 2475
773 ProTyrAspThrLeuValPheAspTyrGluGlySerGlySerAspAlaAlaSerLeu 792
2476 CTTTATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2535
793 SerSerLeuThrSerSerAlaSerAspGlnAspGlnAspTyrAspTyrLeuAsnGluTyr 812
2536 AGCTCCCTGAACTCTCTCAGAGTCAGACAAAGACAGGAGCTATGACTACTTGAACGAATGG 2595
813 GlySerArgPheLysLysLeuAlaAspMetTyrGlyGlyGlyGluAspAsp 829
2596 GGCATTCGCTTCAAGAGCTGCTGACATGTACGAGGCGCGGAGGAGCAGC 2646

RESULT 14
AB025106 2807 bp mRNA linear PRI 07-JAN-2000
LOCUS Homo sapiens mRNA for E-cadherin, complete cds.
ACCESSION AB025106
VERSION AB025106.1 GI:6682962
KEYWORDS E-cadherin.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2807)
AUTHORS Shibamoto,S. and Fukudome,Y.
TITLE E-cadherin mutant
JOURNAL Published Only in Database (2000)
REFERENCE 2 (bases 1 to 2807)
AUTHORS Shibamoto,S. and Fukudome,Y.
TITLE Direct Submission
JOURNAL Submitted (19-MAR-1999) Sayumi Shibamoto, Setsunan University,
Faculty of Pharmaceutical Sciences, Department of Biochemistry, 45-1
Nagatoge-cho, Hirakata, Osaka 573-0101, Japan
(E-mail: sayumi@pharm.setsunan.ac.jp, Tel:81-720-66-3118,
Fax:81-720-66-3117)

FEATURES
Location/Qualifiers
1..2807
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_line="HSC-45M2"
/cell_type="signet ring cell carcinoma"
65..3770
/codon_start=1
/product="E-cadherin"
/protein_id="BAA88957.1"
/db_xref="GI:6682963"
/translation="MGPSRSLSALLLLQVSWLCOBEPCHPGFPAESVTFVPRR
HLERVLGRVNFEDCTQRTAYSLDTRFKVGTGVTYKRLPHNPQIHFLVYA
WDSYRKSTKVLNTVGHHRPPPHQASVSGIQAEILTFPNSRGLRQRDRVIPP
LSCPNKGGPPKLVQIKSNKKGKVFYSITGGADTPVGVFIETERTGWLKVT
PLDERIATYTLFSAVSSNGNAVEDMEILITVDQDNKPEFTQEVFKSVMEGAL
RQTSWEVTDADDVNNTAAAYTILSQDPELPKDMFTINRNTGTVISVITGLD
RESPTVTLVQAADLQGEGLSTATAVITVDINDNPIFNFTYKQVQVENANVV
ITLKTVDADAPNTPAEVYITLNDGGVFVFTNPVNDGILKTAKDLARCNWHP
HSCYPCPLGDLFEAKQYIILHAVTNVVPFVSLTSTATVTDVLDVNEAPIFVPP
EKRVSDSDFGVGBITSTYTAQEPDTMEQKITRYWRDWTANWLEINDPTGAIFFER
LDREDFHKVNSTYALIATDNGSPVATGTGLILLSDVNDNAPIPEPTIIFCFER
NPKPOVINIDADLPNPTSPTELTGCSANWTIOYNDPTQESIILKPKVALBVDY
KINLKMNQNDQVTTLEVSVCCEGAGVCRKAQPVAGLQIPAILGILGLLALL
LILLLLFLRRRAVVKLEPLVPPEDDTRDNYIIDEGGEGEQDFLSQLHRGLDAR
PEVTRNDVATLMSVPRYLPANPDEIGNFDENLKAADTDPAPPYDSLLVFDYEG
SGSEAAASLSLNSSESDDQDYDLNEWGNRFFKLADMYGGGDD"
1202
/note="57base insertion between exon 9 and 10 of
E-cadherin"

variation
ORIGIN
Alignment Scores: 2,186-173 Length: 2807
Pred. No.: 2400.50 Matches: 483
Score:

```


Qy	101	Ile---	Leu	Arg	Arg	His	Gly	Asp	Trp	Val	Val	Ala	Pro	Leu	Val	Pro	Glu	Asn	119	
Db	513	CCAGGTCT	CAAGA	CAC	AAACG	AGACT	GGGT	TCAT	CCCT	CCCAT	CAGCT	GC	CCG	GAAT	572					
Qy	120	Gly	Lys	Gly	Pro	Pro	Glu	Arg	Leu	Asn	Gln	Leu	Lys	Ser	Asn	Lys	Asp	Arg	Thr	139
Db	573	GAATAAGG	TGAAT	CCCC	AAAG	AACT	TGTT	CAGAT	CAAT	CCAC	AGG	CA	CAAG	GA	CA	AA	GA	CA	CA	632
Qy	140	Lys	Leu	Phe	Thr	Ser	Ile	Thr	Gly	Pro	Gly	Ala	Asp	Ser	Pro	Pro	Glu	Gly	Val	159
Db	633	AAGGT	TTTTCT	CAG	CAT	CAC	CGG	CAAG	AGCT	CA	AAAC	CC	CCG	T	TTT	CGG	T	TTT	CGG	692
Qy	160	Val	Glu	Lys	Glu	Thr	Gly	Trp	Leu	Leu	Asn	Lys	Pro	Leu	Asp	Arg	Glu	Leu	Ala	179
Db	693	ATTGAG	AGG	AG	AG	CA	GGT	CGCT	GNA	AGT	GC	AC	AG	CC	TCT	GGAT	AG	AA	GC	752
Qy	180	Lys	Trp	Glu	Leu	Phe	Gly	His	Ala	Val	Ser	Glu	Asn	Gly	Ala	Ser	Val	Glu	Asp	199
Db	753	AAGTAC	AT	CC	TCT	AT	CT	CA	CGC	GT	GC	AT	CA	AT	GGG	AA	GG	CG	GG	812
Qy	200	Asn	Ile	Ser	Ile	Ile	Val	Thr	Asp	Gln	Asn	Asp	His	Lys	Pro	Lys	Phe	Thr	Gln	219
Db	813	CAGAT	AGT	GAT	CA	CAGT	GAC	AGAT	CAG	AT	GAC	AA	CGC	CA	AGT	TT	AC	CC	AG	272
Qy	220	Phe	Arg	Ser	Val	Leu	Glu	Gly	Val	Leu	Pro	Gly	Thr	Ser	Val	Met	Gln	Val	Thr	239
Db	873	TTTGAGG	GAT	CCG	T	TG	CA	AGG	CGC	T	GT	TC	AG	AA	CC	T	CCG	T	GA	292
Qy	240	Thr	Asp	Glu	Asp	Ala	Ile	Thr	Thr	Trp	Asn	Gly	Val	Val	Ala	Thr	Ser	Ile	His	259
Db	933	ACC	GAT	GC	AG	AT	GAT	GC	AT	CA	CC	TAC	AA	CGC	T	GC	AT	CC	AT	312
Qy	260	Gln	Glu	Pro	Lys	Asp	Pro	His	Asp	Leu	Met	Phe	Thr	Ile	His	Arg	Ser	Thr	Gly	279
Db	993	CAG	GAT	CT	GC	AT	GC	CT	CA	AAAA	CA	TG	TT	CA	CT	GA	AT	GG	CA	332
Qy	280	Ser	Val	Ile	Ser	Ser	Gly	Leu	Asp	Arg	Glu	Lys	Val	Pro	Glu	Trp	Thr	Leu	Thr	399
Db	1053	AGT	GT	CT	CA	CT	CT	GG	CT	GG	AC	GG	AG	AT	TAC	CT	CA	TAC	AT	452
Qy	300	Ala	Thr	Asp	Met	Asp	Gly	Asp	Gly	Ser	Thr	Thr	Thr	Ala	Val	Ala	Val	Glu	Ile	319
Db	1113	GCT	GT	CAC	CT	T	CA	GGG	CA	AGG	CT	T	GA	CA	CA	CA	AG	CT	GT	372
Qy	320	Asp	Ala	Asn	Asp	Asn	Ala	Pro	Met	Phe	Asp	Pro	Gln	Lys	Trp	Glu	Ala	His	Val	339
Db	1173	GAT	ATT	AT	GA	CA	ACG	T	CT	CT	CA	CC	CG	AG	CA	GT	AT	CA	GG	392
Qy	340	Asn	Ala	Val	Gly	His	Glu	Val	Gln	Arg	Leu	Thr	Val	Thr	Asp	Leu	Asp	Ala	Pro	359
Db	1233	AAT	GAG	T	CA	AT	G	CC	GC	AT	CG	CA	CT	CA	AA	GT	GC	AT	GA	412
Qy	360	Pro	Ala	Trp	Arg	Ala	Thr	Trp	Leu	Ile	Met	Gly	Lys	Asp	Asp	Gly	Asp	His	Phe	379
Db	1293	CCG	CG	T	GG	AA	AG	T	GT	GT	AC	CG	TA	GC	---	AAC	GA	T	CC	432
Qy	380	Thr	Thr	His	Pro	Glu	Ser	Asn	Gln	Gly	Ile	Leu	Thr	Thr	Arg	Lys	Gly	Leu	Asp	399
Db	1350	GTC	AC	AG	AC															

Qy	460	CysValTyrThrAlaGluAspProAspLys---GluAsnGlnLysIleSerTyrArgile	478
Db	1590	ACATCTTATACCGCTCGAGAGCCGGACACGTTATGTGATCAGAGAATCATCGTATCGGATT	1649
Qy	479	LeuArgAspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaVal	498
Db	1850	TGGAGGACACTGCCAAGTGGTGGAGATTAAACCCAGAGACTGGTGGCAATTTTCCACGGC	1709
Qy	499	GlyThrLeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetVal	518
Db	1710	GCTGAGATGGACAGAGAAGACGCTGAGCATGTGTAAGAACACGACATATGTAGTCTTCATC	1769
Qy	519	LeuAlaMetAspAsnGlySerProProThrThrClyThrGlyThrLeuLeuLeuThrLeu	538
Db	1770	ATGCCACAGATGATGTTCACCCATTTGCCACTGGCACGGGCACTTCTTCCCTGGTCTGT	1829
Qy	539	IleAspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGln	558
Db	1830	TTAGACGTCAATGACAACGCTCCCATCCAGAACCTCGAAACATGCAGTGTCTGCCAGAGG	1889
Qy	559	SerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerPro	578
Db	1890	AACCCACAGCCTCATATCATCATCATCTTGGATCCAGACCTCCCCCCACACAGCTCCCC	1949
Qy	579	PheGlnAlaGlnLeuThrAspAspSerAspIleTyrTriThrAlaGluValAsnGlu	598
Db	1950	TTTACTGCTGAGCTAACCCATGGGGCAGCGTCAACTGHCACCTTGAATATATGACGCA	2009
Qy	599	Gly---AspThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspVal	617
Db	2010	GCTCAAGAAATCTCTCATTTTTTGCAACCAAGAAAGAGACTTAGAGATTGGCGAATACAAAT	2069
Qy	618	HisLeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrVal	637
Db	2070	CATCTCAAGCTCGGGATATACCAGACAAAGACCAGGTGACCAGTTGGACGTCCATGTG	2129
Qy	638	CysAspCysHisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPhe----	655
Db	2130	TGTGACTGTGAAGGACCGGTCAACAACCTGC-----ATGAAGCGGGAATCGTGGCA	2180
Qy	656	-----IleLeuProValLeuGlyAlaValLeuAlaLeuLeuPhe	668
Db	2181	GCAGGATTGAAGTCTCTGCCATCTCGGAATCTTTGGAGGATCTCTGCCCTGCTGATT	2240
Qy	669	LeuLeuValLeuLeuLeuValArgLysLysArgLysIleLysGluProLeuLeu	688
Db	2241	CTGATCTGCTGCTCTACTGTTTCTACGGAGAGAACGGTGTCTAAAGAGCCCTGCTG	2300
Qy	689	LeuProGluAspAspThrArgAspAsnValPheTyrTyrClyGluGluGlyGlyGlu	708
Db	2301	CCACCAGATGATATACCCGGGCAATGTGTATTCTATGATGAAGAAGAGAGTGGAGAA	2360
Qy	709	GluAspGlnAspTyrAspIleThrGlnLeuHisArgGlyLeuGluAlaArgProGluVal	728
Db	2361	GAAGACCAGACATTGATTTGAGCCAGCTGCACAGGGGCTTGGATCCCCACCGAAGTG	2420
Qy	729	ValLeuArgAsnAspValAlaProThrIleLeuProThrProMetTyrArgProArgPro	748
Db	2421	ACT---CGAAATGATGTGGTCCACCCCTCATGAGGGGTGCCCAAGTATCGTCCCGCTCT	2477
Qy	749	AlaAsnProAspGluIleGlyAsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAsp	768
Db	2478	GCCATCTGATGAAATTTGAAACTTCAUCGATGAAACCTGHAAGCAGCCGACAGCGCAC	2537
Qy	769	ProThrAlaProTyrAspThrLeuLeuValPheAspTyrClyGlySerGlySerAsp	788
Db	2538	CCACGCGCACCCCTTACGACTCTCTGTGGTGTTCGATTACGAGGCGAGTGGTCTTGAA	2597
Qy	789	AlaAlaSerLeuSerSerLeuThrSerSerAlaSerAspGlnAspGlnAspTyrAspTyr	808
Db	2598	CCCGCTAGCTGAGCTCACTGAATCTCTCTGAGTCGATCAGACACGAGACTACGATTAT	2657
Qy	809	LeuAsnGluTyrPglySerArgPheLysLysLeuAlaAspMetTyrGlyGlyGlyGluAsp	828

Db 2658 CTGAACGAGTGGGCAACCGATTCAAGAGCTGGCGGACATGTACGGCGGTGGCGAGGAC 2717
Qy 829 Asp 829
Db 2718 GAC 2720

Search completed: September 23, 2004, 00:00:57
Job time : 8254 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 22, 2004, 20:38:13 ; Search time 766 Seconds

(without alignments)
4597.593 Million cell updates/sec

Title: US-09-916-849a-1

Perfect score: 4369

Sequence: 1 MGUPRGPPLASLLLLQVCWLQ.....NEWGSRFKKLADNYGGGDD 829

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2 1/USPTO.spool.h/US09916849/runat 21092004 145802 7002/app query.fasta_1.967
-DB=N Geneseq 29Jan04 -OPT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_MIN=100 -THR_MAX=100 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -NAPLEN=2000000000
-USER=US09916849 @CGN 1.1 513 @runat 21092004 145802 7002 -ICPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4369	100.0	3171	6 ABK92214	Abk92214 Prostate
2	4369	100.0	3171	6 ABN89393	Abn89393 Human P-c
3	4369	100.0	3171	6 ABT07755	Abt07755 Breast ca
4	4369	100.0	3171	6 ABK52038	Abk52038 DNA encod
5	4369	100.0	3171	6 ABN97264	Abn97264 Gene #376
6	4369	100.0	3171	7 ACAS6750	Acas6750 Human sig
7	4369	100.0	3171	7 ABZ24736	Abz24736 Human P-C
8	4369	100.0	3171	9 ADD14795	Add14795 Human src

9	4369	100.0	3777	8 ACH03801	Ach03801 Human CDN
10	4361	99.8	3205	6 ABQ81547	Abq81547 Gene up-r
11	4361	99.8	3205	7 ACC72821	Acc72821 Human can
12	4361	99.8	3205	7 ABX76155	Abx76155 Lung can
13	4361	99.8	3205	7 ABX76399	Abx76399 Lung can
14	4348	99.5	3219	6 AAD27637	Aad27637 Human col
15	2449.5	56.1	4333	2 AAQ11563	Aaq11563 Encodes E
16	2420	55.4	4778	2 AAX56022	Aax56022 Wild-type
17	2420	55.4	4778	4 AAF86887	Aaf86887 Human E-c
18	2420	55.4	4828	6 ABS76385	Abs76385 CDNA enco
19	2420	55.4	4828	6 ABV94337	Abv94337 Breast ca
20	2388	54.7	2768	6 ABN89352	Abn89352 Mouse E-c
21	2388	54.7	2768	6 ABN89353	Abn89353 Mouse E-c
22	2388	54.7	4396	7 ABT42198	Abt42198 Toxicity
23	2388	54.7	4396	9 ADB53244	Adb53244 Primary r
24	2349.5	53.8	2808	2 AAT05764	Aat05764 Human E-c
25	2342.5	53.6	2808	6 ABN89392	Abn89392 Human E-c
26	2342.5	53.6	2815	2 AAQ65487	Aaq65487 Sequence
27	2311	52.9	2156	2 AAQ11562	Aaq11562 Partial e
28	1713	39.2	3427	7 ABX63589	Abx63589 Human CDN
29	1713	39.2	3427	8 ADA24495	Ada24495 Human CDN
30	1713	39.2	4122	7 ACC72793	Acc72793 Human can
31	1708	39.1	4136	5 AAS78503	Aas78503 DNA encod
32	1707	39.1	2853	6 ABN89391	Abn89391 Human N-c
33	1707	39.1	2853	7 ACC50097	Acc50097 Breast ca
34	1704.5	39.0	2824	6 ABN89350	Abn89350 Chicken N
35	1704.5	39.0	2824	6 ABN89351	Abn89351 Chicken N
36	1701.5	38.9	4350	9 ADB58603	Adb58603 Toxicity-
37	1701.5	38.9	4350	9 ADB53243	Adb53243 Primary r
38	1697.5	38.9	4321	6 ABT99593	Abt99593 Mouse isc
39	1684.5	38.6	3875	2 AAQ11561	Aaq11561 Partial e
40	1662	38.0	3448	9 ADD14630	Add14630 Human src
41	1641	37.6	3448	6 ABZ35199	Abz35199 Human gen
42	1638	37.5	3310	9 ADB62391	Adb62391 Human CDN
43	1633.5	37.4	3451	9 ADD12666	Add12666 Human CDN
44	1621.5	37.1	3048	2 AAT85433	Aat85433 Human cad
45	1621.5	37.1	3048	2 AAT61920	Aat61920 Full leng

ALIGNMENTS

RESULT 1

ABK92214

ID ABK92214 standard; DNA; 3171 BP.

XX

AC ABK92214;

XX

DT 15-AUG-2002 (first entry)

XX

DE Prostate cancer-associated DNA sequence #100.

XX

KW Prostate cancer; prostate tumour tissue; human; mammal; cytostatic;

XX Gene therapy; gene; ds.

OS Mammalia.

XX

PN WO200230268-A2.

XX

PD 18-APR-2002.

XX

PF 12-OCT-2001; 2001WO-US032045.

XX

PR 13-OCT-2000; 2000US-00687576.

PR

PR 08-DEC-2000; 2000US-0073288.

PR

PR 08-DEC-2000; 2000US-00733742.

PR

PR 24-JAN-2001; 2001US-02639576.

PR

PR 16-MAR-2001; 2001US-0276791P.

PR

PR 16-MAR-2001; 2001US-0276888P.

PR

PR 06-APR-2001; 2001US-0281922P.

PR

PR 24-APR-2001; 2001US-0286214P.

PR

PR 30-APR-2001; 2001US-00847046.

PR

PR 04-MAY-2001; 2001US-0288589P.

XX

QY 21 CysAlaIaSerGluProCysArgAlaValPheArgGluAlaValThrLeuGluAla 40
DB 114 TGC GCGGCTCCGAGCGCGTGC GCGGCGGCTCTTCAGGAGGCTGAAGTGCCTTGGAGGCG 173
QY 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
DB 174 GGAGCGCGGAGCAGGAGCGCGCGCCAGCGCTGGGGAAAGTATTCAATGCGCTGCCCTGGG 233
QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
DB 234 CAAGAGCCAGCTCTGTTAGCACTGATAATGATGACTTCACTGTGCGGAATGGCAGACA 293
QY 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
DB 294 GTCCAGGAAGAAGGTCACTGAAGAAAGGATCCATTTGAAGATCTTCCCATCCAAACGT 353
QY 101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
DB 354 ATCTTACGACACACAGAGAGATTGGTGGTGTCTCCAAATATCTGTCCCTGAAATGGC 413
QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAsnThrLys 140
DB 414 AAGGGTCCCTTCCCGCAGAGACTGAATCAGCTCAAGTCTAATAAGATAGAGACCAAG 473
QY 141 IlePheTrpSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal 160
DB 474 ATTTTCTACACATCACGGGCGCGGGGCGAGACGCCCTCGAGGGTGTCTTCGCTGTA 533
QY 161 GluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
DB 534 GAGAAGGAGACAGGCTGTTGTTCTTGAATAAGCCACTGGACCGGGAGGAGATTGCCAAG 593
QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
DB 594 TATGAGCTTTTGGCCACCTGTGTACAGNATGGTGCCTCAGTGGAGGACCCCATGAAC 653
QY 201 IleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
DB 654 ATCTCCATCATCGTACCGACCAAGATACCCAAAGCCCAAGTTTACCCAGGACACCTTC 713
QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
DB 714 CGAGGAGTGCTTAGAGGAGTCTACAGGACTTCTGTGATGACAGTGCAGCCACA 773
QY 241 AspGluAspAspAlaIleTyrThrTrpAsnGlyValValAlaTrpSerIleHisSerGln 260
DB 774 GATGAGGATGATGCCATCACTACACTACAAATGGGTGGTGTGTTACTCCATCCATAGCAA 833
QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
DB 834 GAACCAAGAGACCCACAGCCTCATGTTCATCAATCCCGAGCAGGACCCATCAGC 893
QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
DB 894 GTCATCTCCAGTGCCTGACCGGGAAAAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
QY 301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValGluIleLeuAsp 320
DB 954 ACAGACATGGATGGGAGCGGCTCCACCACCGCAGTGGCAGTAGTGGAGATCTTTGAT 1013
QY 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
DB 1014 GCCAATGACATGTCTCCATGTTTGAACCCCAAGATACGAGGCCCATGTGCTGAGAAT 1073
QY 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
DB 1074 GCAGTGGGCCATGAGTGCAGAGGCTACCGGTCACTGTCTGGACGCGCCCACTCAACA 1133
QY 361 AlaTrpArgAlaThrTyrLeuIleMetGlyLysAspGlyAspHisPheThrIleThr 380
DB 1134 CGTGGCGTGCACCTACTTATCATGGGCGGTGACGCGGGACCAATTTTACCATCACC 1193
QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400

DB 1194 ACCCACTCTGAGAGCAACAGGGCATCTGACACCAGNAGGTTTGGATTTTGGAGGCC 1253
QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
DB 1254 AAAAACCCAGCACACCCTGTAGCTTGAAGTACCACAGAGGCCCTTTTGTGTGAAGCTC 1313
QY 421 ProThrSerThrAlaThrIleValIleHisValGluAspValAsnGluAlaProValPhe 440
DB 1314 CCAACTCCACAGCCACCATAGTGTCCAGTGGAGATGTGAATGAGGCACCTGTGTGT 1373
QY 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
DB 1374 GTCCCACTCCAAAGTGGTTGAGTCCAGAGGGCATCCCACTGGGGAGGCTGTGTGT 1433
QY 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
DB 1434 GTCTACACTGCAGAAAGACCTGACAAAGAGATCAAAAGATCAGTACCCTCTCTGAGA 1493
QY 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
DB 1494 GACCCAGCAGGGTGGCTAGCCATGCCACAGCAGTGGGCAGGTCAACAGTGTGGCACC 1553
QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
DB 1554 CTGCACTCTGAGGATGAGCAGTTTGTGAGGAAACAACATCTATGAAGTCACTGGTCTTGCC 1613
QY 521 MetAspAsnGlySerProProThrThrGlyThrLeuLeuLeuThrLeuIleAsp 540
DB 1614 ATGGCATGTGAAGCCCTCCCACTGGCAGCGAACCCTTCTGCTAACACTGATTGAT 1673
QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
DB 1674 GTCAACGACCATGGCCAGTCCCTGAGCCCGTGCAGATCACCATCTGCAACCAAGCCCT 1733
QY 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
DB 1734 GTGCGCCAGTGTCTGAACATCAGCAGAGGACCTGTCTCCCACTCCCTCCCTTCCAG 1793
QY 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGluGlyAsp 600
DB 1794 GCCAGCTCACAGATGACTCAGACATCTACTGACCGCAGAGGTCAACAGAGAAAGTGAC 1853
QY 601 ThrValValLeuSerLeuLysIlePheLeuLysGlnAspThrTyrAspValHisLeuSer 620
DB 1854 ACAGTGGTGTCTGCTGAAAGTTCCTGAAGCAGGATACATATGAGCTGCACCTTCT 1913
QY 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
DB 1914 CTGCTGACCATGGCAACAAGAGCAGCTGACGGTGTACAGGGCCACTGTGTGGACTGC 1973
QY 641 HisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValLeu 660
DB 1974 CATGGCCATGTGAAACCTGCCCTCGAACCTGGAAAGAGGTTTCATCCTCCCTGTGTGTG 2033
QY 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuValArgLysLys 680
DB 2034 GGGGCTGTCTGTGCTGTCTGTCTCTGTGTGTGTGTCTTGTGTGTGTGTGTGTGTGTGT 2093
QY 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
DB 2094 CGGAAGATCAAGGAGCCCTCTCTACTCCAGAAAGATGACACCGCTGACCAAGCTCTTCTAC 2153
QY 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
DB 2154 TATGGCAAGAGGGGGTGGGAGAGGACCAAGGACTATGACATCACCAGCTCCACCGA 2213
QY 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
DB 2214 GTCTGGAGGCCAGCGCGAGGTGTCTCTCGCAATGACGTGGCACCACCAACCATCTCCG 2273
QY 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760

Db 2274 ACACCCATGTACCGTCTAGCCAGCCCAACCCAGATGAAATCGGCAACTTTATAATTGAG 2333
 Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
 Db 2334 AACCTGAAGCGGTACACAGACCCACAGCCCGCCCTACGACACCCCTTGGTGTTC 2393
 Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
 Db 2394 GACTATGAGGCGAGCGGCTCCGAGCGCGCTGAGCTCCCTCAGCTCCCTCAGCTCCGCTCC 2453
 Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLeuLysLeuAla 820
 Db 2454 GACCAAGACCAAGATTACGATTATCTGAACAGTGGCGGACCGCTTCAAGAGCTGGCA 2513
 Qy 821 AspMetTyrGlyGlyGluAspAsp 829
 Db 2514 GACATGTACGGTGGCGGAGGAGCAG 2540

RESULT 3
 ABT07755
 ID ABT07755 standard; DNA; 3171 BP.
 XX AC
 AC ABT07755;
 XX XX
 DT 14-NOV-2002 (first entry)
 XX XX
 DE Breast cancer-associated gene sequence 63.
 XX XX
 KW Gene; ds; breast cancer; breast cancer-associated gene sequence;
 KW drug development; pharmacogenetics; biosensor development.
 XX XX
 OS Unidentified.
 XX XX
 FN WO200259377-A2.
 XX XX
 PD 01-AUG-2002.
 XX XX
 PF 24-JAN-2002; 2002WO-US002242.
 XX XX
 PR 24-JAN-2001; 2001US-0263965P.
 PR 02-FEB-2001; 2001US-0265928P.
 PR 09-APR-2001; 2001US-00829472.
 PR 09-APR-2001; 2001US-0282698P.
 PR 04-MAY-2001; 2001US-0288590P.
 PR 29-MAY-2001; 2001US-029443P.
 XX XX
 FA (BOSB-) EOS BIOTECHNOLOGY INC.
 XX XX
 PI Mack DH, Gish KC, Afar D;
 XX XX
 DR N-PSDB; ABJ05598.
 XX XX
 WPI: 2002-583738/62.
 XX XX
 Detecting a breast cancer-associated transcript in a patient's cell,
 useful for diagnosing breast cancer, comprises contacting a biological
 sample with a polynucleotide that selectively hybridizes with breast
 cancer nucleic acids.
 XX XX
 Claim 9; Page 404-405; 414pp; English.
 XX XX
 The invention comprises a method of detecting a breast cancer-associated
 transcript in a cell from a patient. The method of the invention involves
 contacting a biological sample from the patient with a nucleotide that
 hybridizes to one of the 69 breast cancer-associated gene sequences shown
 in the specification. The method of the invention is useful in the
 diagnosis or prognosis of breast cancer, and for detecting genes that are
 up or down-regulated in breast cancer cells. Genes identified by the
 method of the invention can be used in diagnostic purposes and also as
 targets for screening for therapeutic compounds that modulate breast
 cancer (e.g. hormones or antibodies). Identification of genes that are
 over or under expressed in breast cancer can additionally provide high-
 resolution, high-sensitivity datasets which can be used in the areas of
 diagnostics, therapeutics, drug development, pharmacogenetics, protein

CC structure and biosensor development. DNA sequences ABT07693 - ABT07761
 CC represent the 69 breast cancer-associated gene sequences of the invention
 XX Sequence 3171 BP; 740 A; 903 C; 864 G; 664 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 0 Length: 3171
 Score: 4369.00 Matches: 829
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-916-849A-1 (1-829) x ABT07755 (1-3171)

Qy 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln 20
 Db 54 ATGGGGCTCCCTCGTGACCTCTCGCGTCTCTCCCTCTCCAGGTTTGGCTGGCTGAG 113
 Qy 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
 Db 114 TCGCGCGCTCCGAGCGCTGCCGGCGGTCTTTCAGGAGGCTGAAAGTACCTTGGAGGG 173
 Qy 41 GlyGlyValaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
 Db 174 GGAGGGCCGAGAGGAGGCGCCGCGAGCGCTGGGGAAAGTATTTCATGGGCTGCCCTGG 233
 Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
 Db 234 CAAGAGCCAGCTCTGTTTAGCACTGATAATGATGACTTCACTGTGCGGAATGCGGAGCA 293
 Qy 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
 Db 294 GTCCAGGAAGAAGGTCACTGAAGGAAGAAGAAATCCATTGAAGATCTTCCATCCAAACGT 353
 Qy 101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
 Db 354 ATCTTACGAAGACACACAAGAGAGATTGGGTGGTGTCTCCCAATATCTGTCCCTGAAATG 413
 Qy 121 LysGlyProPheProGlnhArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
 Db 414 AAGGGTCCCTTCCCCCAGAGACTGAATCAGCTCAAGTCTTAATAAAGATAGAGACACCAAG 473
 Qy 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal 160
 Db 474 ATTTCTACAGCATCACGGGGCGGGGGGAGAGACAGCCCTTCTGAGGGTGTCTTCGCTGA 533
 Qy 161 GluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluLeuLys 180
 Db 534 GAGAGGAGACAGCGCTGTTGTTGTAATAAGCCACTGGACCGGAGAGATTGCCAAG 593
 Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
 Db 594 TATGAGCTTTTGGCCACCGCTGTGCAGAGAATGGTCCCTCAGTGAGAGACCCCAAGAAC 653
 Qy 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
 Db 654 ATCTCCATCATCGTACCGACCAAGTACCAAGCCCAAGCCCAAGTTTACCAGGACACCTTC 713
 Qy 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
 Db 714 CGAGGGAGGTCTTGAAGGGAGTCTTACAGGAGTCTCTGTGATCAGGTGAGGACCA 773
 Qy 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
 Db 774 GATGAGATGATGCCATCTACACCTACATGATGGGTGGTGTCTTACTCCATCCATAGCCA 833
 Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
 Db 834 GAACCAAAAGAGCCCAACACACCTCATGTTTCACAATTACCGGAGCAGACGACCATCAGC 893
 Qy 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300

Db 1674 GTCAACGACCATGGCCAGTCCCTGAGCCCGCTCAGATCACCATCTCCAAACCAAGCCCT 1733
 Qy 561 ValArgHisValLeuAenIleThrAspIysAspLeuSerProHisThrSerProPheGln 580
 Db 1734 GTGGCCACGCTGCTGAACATCAGCAGCAAGGACCTGTCTCCCAACACCTCCCTTCCAG 1793
 Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpAlaGluValAsnGluGluGlyAsp 600
 Db 1794 CCCAGCTCACAGATGACTCAGACATCTACTGAGCGCAGAGGTCAACGAGGAGGTGAC 1853
 Qy 601 ThrValValLeuSerLeuIysPheLeuIysGlnAspThrTyrAspValHisLeuSer 620
 Db 1854 ACAGTGGTCTTGTCCCTGAAGAAGTCTCTGAAGCAGGATACATATGACGTGCACCTTTCT 1913
 Qy 621 LeuSerAspHisGlyAsnLysGluLeuThrValIleArgAlaThrValCysAspCys 640
 Db 1914 CTGTCTGACCATGGCAACAAAGACAGCTGACGGTGATCAGGGCCACTGTGTGGACTGC 1973
 Qy 641 HisGlyHisValGluThrCysProGlyProTyrPheGlyGlyPheIleLeuProValLeu 660
 Db 1974 CATGCCCATGTGAAACCTGACCTGGACCTCGGAAGAGGTTTCACTCCCTGTGCTG 2033
 Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuValArgLysLys 680
 Db 2034 GGGGCTGTCTGGCTCTGCT 2093
 Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
 Db 2094 CGGAGATCAAGAGGCCCTCTACTCCAGAGATGACACCCGCTGACCAAGCTTCTAC 2153
 Qy 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
 Db 2154 TATGCGAAGAGGGGGTGGGAAAGAGGACCAAGCACTATGACATCACCAGCTCCACCGA 2213
 Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
 Db 2214 GGTCTGGAGGCGAGCGGAGTGTCTCTCGCAATGACGTGGCACCACCAATCATCCG 2273
 Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
 Db 2274 ACACCATGTACCGTCTTAGCCAGCAACCCAGATGAAATCGGCACACTTATAATAG 2333
 Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProTyrAspThrLeuLeuValPhe 780
 Db 2334 AACCTGAAGGGCGCTTAAACACAGACCCCAACGCCCCCTTACGACACCCCTTCTGGTTC 2393
 Qy 781 AspTyrGluGlySerGlySerAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
 Db 2394 GACTATGAGGCGAGCGGCTCCGACGCGCTCCCTGAGCTCCCTCCTCCTCCGCTCC 2453
 Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluThrGlySerArgPheLysLysLeuAla 820
 Db 2454 GACCAAGACCAAGATTACGATTATCTGAACAGTGGGCGAGCCGCTTCAAGAGCTGGCA 2513
 Qy 821 AspMetTyrGlyGlyGluAspAsp 829
 Db 2514 GACATGTACGTGGCGGGGAGGAGCAGC 2540
 RESULT 5
 ID ABN97264
 XX ABN97264 standard; DNA; 3171 BP.
 AC ABN97264;
 XX
 DT 13-AUG-2002 (first entry)
 DE Gene #3762 used to diagnose liver cancer.
 DE
 KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotrophic;
 KW metastatic liver tumor; cytostatic; expression profile; disease state;
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.
 XX Homo sapiens.
 OS

XX PN WO200229103-A2.
 XX PD 11-APR-2002.
 XX PF 02-OCT-2001; 2001WO-030589.
 XX PR 02-OCT-2000; 2000US-0237054P.
 XX (GENE-) GENE LOGIC INC.
 XX PA
 XX PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
 XX MPI; 2002-426119/45.
 XX DR
 XX DX
 PT Diagnosing and detecting the progression of liver cancer, hepatocellular
 PT carcinoma or metastatic liver tumor in a patient, involves detecting the
 PT level of expression of two or more genes in a liver tissue sample.
 XX
 PS Claim 1; SEQ ID NO 3762; 298pp; English.
 CC The invention relates to a novel method for diagnosing and detecting the
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
 CC tumor in a patient, and differentiating metastatic liver cancer from
 CC hepatocellular carcinoma in a patient, involving detecting the level of
 CC expression of two or more genes represented in ABN93503-ABN97455 in a
 CC tissue sample. The method of the invention has hepatotrophic, and
 CC cytostatic activity. The method is useful for diagnosing and detecting
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic
 CC liver carcinoma in a patient. The method is useful for identifying
 CC expression profiles which serve as useful diagnostic markers as well as
 CC markers that can be used to monitor disease states, disease progression,
 CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 3171 BP; 740 A; 903 C; 864 G; 564 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0 Length: 3171
 Score: 4369.00 Matches: 829
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0

US-09-916-849A-1 (1-829) x ABN97264 (1-3171)

Qy 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuLeuValCysTrpLeuGln 20
 Db 54 ATGGGGCTCCCTCGTGGACCTCTCGGCTCTCTCCCTCTCCAGTTTGTGGCTGCAG 113
 Qy 21 CysAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
 Db 114 TGGCGGCGCTCCGAGCGGTGCGCGGCTTTCAGGAGGCTGAAGTACCTTGGAGCGG 173
 Qy 41 GlyGlyValAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
 Db 174 GGAGCGCGGAGCAGAGCGCCGCGCGCTGGGAAAGTATTTCATGGGCTGCCCTGGG 233
 Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspAspPheThrValArgAsnGlyCysThr 80
 Db 234 CAAGAGCCAGCTCTGTGTTAGCACTGATATATGATGACTTCACTGTCCGGAATGGCGAGACA 293
 Qy 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
 Db 294 GTCCAGAAAGAAAGGTCACTGAGGAAAGANTCCATTGAGATCTTCCCATCCAAAGCT 353
 Qy 101 IleLeuArgArgHisGlyArgAspTrpValValAlaProIleSerValProGluAsnGly 120
 Db 354 ATCTACGAAGACACAAAGAGAGATTGGGTGGTGTCTCCAAATATCTGTCTCCGTAAGTGGC 413

QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
DB 414 AAGGTCCTCCCTCCCCAGAGATGAATCAGCTCAAGTCTAATAAAGATAGAGACCAAG 473
QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
DB 474 ATTTTCTACAGCATCAGGGGCGGGGCGAGACAGCCCCCTCAGGGGTGCTTCGCTGTA 533
QY 161 GlutysGluThrGlyTrpLeuLeuAsnLysPLeuAspArgGluGluIleAlaLys 180
DB 534 GAGAAGGAGAGCAGGCTGTTGTTTGAATAAGCCACTGACCGGGAGGAGATTGCAAG 593
QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
DB 594 TATGAGTCTTTGGCCACGCTGTGTGAGAGATGGTGCCTCAGTGGAGGCCCATGAAC 653
QY 201 IleSerIleLeuValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
DB 654 ATCTCCATCATCGTGCAGCCAGCATGACCAAGCCCAAGTTTACCCAGGACACCTTC 713
QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
DB 714 CGAGGAGTGTCTTAGAGGAGTCTTACCAGGTACTTCTGTGATGCGAGGTGACAGCCACA 773
QY 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
DB 774 GATGAGGATGATGCATCTACACCTACATCAATGGGGTGGTTGCTTACTCCATCCATAGCCAA 833
QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
DB 834 GAACCAAGAGCCACACGACCTCATGTTCCAAATTCACCGGAGCACAGGCACCATCAGC 893
QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
DB 894 GTCATCTCCAGTGGCTGGACCGGGAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
QY 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluIleLeuAsp 320
DB 954 ACAGACATGATGGGACGGCTCCACCACACGCGAGTGCAGTGTGGAGATCCTTGAT 1013
QY 321 AlaAsnAspAlaProMetPheAspProGluLysTyrGluAlaHisValProGluAsn 340
DB 1014 GCCATGACATGCTCCATGTTTGACCCCGAAGTACGAGGCCCATGTGCTTGAGAT 1073
QY 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
DB 1074 GCAGTGGGCCATGAGGTGCAGAGGCTGACGCTCACTGTCGACGCCCCCACTCACCA 1133
QY 361 AlaTrpArgAlaThrTyrLeuIleMetGlyAspAspGlyAspHisPheThrIleThr 380
DB 1134 GCGTGGCGTCCACCTACCTTATCATGGCGGTGACGACGGGACCATTTTACCATCAC 1193
QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
DB 1194 ACCACCTCTGAGACACACCGGCATCTCGACACCGAGGGTTTGGATTTTGGGCC 1253
QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
DB 1254 AAAAAACGACACACCTGTACGTGAAGTACCAACAGAGGCCCTTTTGTGCTGAAGCTC 1313
QY 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
DB 1314 CCACCTCCACAGCCACATAGTGGTCCACGTGAGGATGTGAATGAGGCACCTGTGTTT 1373
QY 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
DB 1374 GTCCACCTCCAAAGTCGTTGAGTCCAGAGGGCATCCCCACTGGGAGCCCTGTGTGT 1433
QY 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
DB 1434 GTCTACACTGCAGAGACCTGTGACAGGAGAACTCAAAGATCAGCTACCGCATCCTGTGAG 1493
QY 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500

DB 1494 GACCCAGCAGGGTGGCTAGCCATGACCCAGACAGTGGGCAGGTACAGCTGTGGGCACC 1553
QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
DB 1554 CTCGACCGTGAAGATGAGCAGTTCGTGAGGAACAACATCTATGAAGTCATGTGCTGGCC 1613
QY 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
DB 1614 ATGGCAATGGGAAGCCCTCCCACTGGCACCGGAACCTTCTGCTTAACACTGATTGAT 1673
QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
DB 1674 GTCAACGACCATGGCCAGTCCCTGAGCCCCGTGAGATCACCATCTGCAACCAAGCCCT 1733
QY 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
DB 1734 GTGCGCCACGTCTGAACATCACGGACAAGGACCTGTCTCCCAACACCTCCCTTTCCAG 1793
QY 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGluGlyAsp 600
DB 1794 GCCCAGCTCACAGATGACTCAGACATCTCTGGACGGCAGAGGTCAACGAGGAAGTAC 1853
QY 601 ThrValValLeuSerLeuLysLysPheLysGlnAspThrTyrAspValHisLeuSer 620
DB 1854 ACAGTGTCTTCTGCTGAAAGAAGTTCCTGAAGCAGGATACATATGACGTGCACCTTCT 1913
QY 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
DB 1914 CTGTCTGACCATGGCAACAAGAGCAGCTGACGGTGATCAGGGCCACTGTGTGCACCTGC 1973
QY 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu 660
DB 1974 CATGGCCATGTGAAACCTGCCCTGAGCCCTGGAAAGGAGGTTTCATCTCCTCCTGCTG 2033
QY 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValLargLysLys 680
DB 2034 GGGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2093
QY 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
DB 2094 CCGAGATCAAGAGCCCTCTCTACTCCAGNAGATGACCCCGTGACACGCTCTTCTAC 2153
QY 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
DB 2154 TATGCGAAGAGGGGGTGGGAGAGGACCAAGGACTATGACATCACCCAGCTCCACCGA 2213
QY 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
DB 2214 GGTCTGGAGGCGCAGGCCGAGGTGTTCTCGCAATGACGTGGCACCAACCATCATCCCG 2273
QY 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
DB 2274 ACACCCATGTACCGTCTTAGGCCACCAACCCAGATGAAATCGGCAACTTTATATTGAG 2333
QY 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
DB 2334 AACCTGAAGGGGGCTTAACACAGACCCCGCCCTACGACACCCCTCTTGTGTGTC 2393
QY 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
DB 2394 GACTATGAGGCGAGGGCTCCGACCGCGCTCCCTGAGCTCCCTCACCTCTCCGCTCC 2453
QY 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLeuAla 820
DB 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCCGCTTCAAGAAGCTGCA 2513
QY 821 AspMetTyrGlyGlyGluAspAsp 829
DB 2514 GACATGTACGTGGCGGGAGGAGCGAC 2540

RESULT 6
ACA56750

XX	AC	ACA56750 standard; cDNA; 3171 BP.	QY	21	CysAlaIaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla	40
XX	AC	ACA56750;	Db	114	TGCGGGGCTCCGAGCGCTGCGGGCGGTCTTCAGGGAGGCTGAAGTACACCTTGAGGCG	173
XX	DT	06-JUN-2003 (first entry)	QY	41	GlyGlyValaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly	60
XX	DE	Human signalling pathway polynucleotide probe SEQ ID NO 1348.	Db	174	GGAGCGGGAGCAGGAGCGCGGCGAGCGCTGGGAAAGTATTTCATGGGCTGCCCTGGG	233
XX	KW	Human; probe; ss; array element; Parkinson's disease;	QY	61	GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr	80
XX	KW	signalling pathway population; cancer; adenocarcinoma; leukaemia;	Db	234	CAAGAGCCAGCTCTGTTAGCACTGATATGATGACTTCACTGTGCGGAATGGCGAGACA	293
XX	KW	immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.	QY	81	ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg	100
XX	OS	Homo sapiens.	Db	294	GTCCAGGAAGAAGGTCACTGAGGAAGAAGATCCATTGAAGATCTCCCATCCAAAGCT	353
XX	FN	US6500938-B1.	QY	101	IleLeuArgArgHisLysArgAspTyrValValaProIleSerValProGluAsnGly	120
XX	PD	31-DEC-2002.	Db	354	ATCTTACGAAGACACAGAGAGATTGGGTGTTGCTCCATATCTGTCCTGAAATGGC	413
XX	PF	30-JAN-1998; 98US-00016434.	QY	121	LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspAspThrLys	140
XX	PR	30-JAN-1998; 98US-00016434.	Db	414	AAGGGTCCCTTCCCCAGAGACTGAATCAGCTCAAGTCTAATAAAGATAGAGACCCAG	473
XX	PA	(INCY-) INCYTE GENOMICS INC.	QY	141	IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal	160
XX	PI	Au-Young J, Seilhamer JJ;	Db	474	ATTTTCTACAGCATCACGGGGCGGGGCGAGACAGCCCCCTGAGGGTGTCTTCGTGTA	533
XX	DR	WPI; 2003-352189/33.	QY	161	GluLysGluThrGlyTyrPheLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys	180
XX	PT	Combination of polynucleotide probes, useful as array elements in a	Db	534	GAGAAGGAGACAGAGGTGGTGTGTTGTTGAATAAGCCACTGGACCGGGAGGAGTTGCCAAG	593
XX	PT	microarray for monitoring the expression of a number of target	QY	181	TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn	200
XX	XX	polynucleotides.	Db	594	TATGAGCTCTTTGGCCAGCTGTGTACAGAAATGGTGCCTCAGTGGAGGACCCCATGAAC	653
XX	PS	Claim 1; SEQ ID NO 1348; 65pp; English.	QY	201	IleSerIlelleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe	220
XX	CC	The invention relates to a combination which, comprises a number of	Db	654	ATCTCCATCATGTCGACCGACAGAAATGACCACAGCCCAAGTTTACCAGGACACCTTC	713
XX	CC	polynucleotide probes comprising a sequence selected from one of the 1490	QY	221	ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr	240
XX	CC	sequences mentioned in the specification. The combination is useful as an	Db	714	CGAGGAGTGCTTAGAGGGAGTCTCTACAGTACTTCTGTGATGCGAGTGCACAGCCACA	773
XX	CC	array element in a microarray for monitoring the expression of a number	QY	241	AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln	260
XX	CC	of target polynucleotides. The microarray is particularly useful in the	Db	774	GATGAGGATGATCCATCTACACCTACAAATGGGTGGTGTGTTACTTCCATCCATAGCCAA	833
XX	CC	diagnosis and treatment of cancer and immunopathology and neuropathology.	QY	261	GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer	280
XX	CC	The microarray is useful in diagnostics and treatment regimens, drug	Db	834	GAACCAAGGAGCCACACAGCTCATGTGTACAAATTCACCGGAGCAGGACCATCAGC	893
XX	CC	discovery and development, toxicological and carcinogenicity studies,	QY	281	ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla	300
XX	CC	forensics and pharmacogenomics. The microarray is also useful for	Db	894	GTCACTCCAGTGGCTGGACCGGAAAAAGTCCCTGAGTACACACTGACCATCCAGGCC	953
XX	CC	monitoring progression of diseases and for developing sophisticated	QY	301	ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValGluIleLeuAsp	320
XX	CC	profiles for the effects of currently available therapeutic drugs. The	Db	954	ACAGATGGATGGGAGCGGTCCACCCAGCGGAGTGGCAGTAGTGGAGATCCTTGAT	1013
XX	CC	combination is also useful for purifying a subpopulation of mRNAs, cDNAs	QY	321	AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn	340
XX	CC	and genomic fragments and in research and diagnostic applications. The	Db	1014	GCATATGACATGCTCCCATGTTTGACCCCCAGAGTACGAGGCCCATGTCCTGAGAAAT	1073
XX	CC	array can detect changes in expression in a large number of genes coding	QY	341	AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro	360
XX	CC	for different signaling pathway populations which can be used to diagnose	Db	1074	GCAGTGGGCCATGAGGTGCAGAGGTGACGGTCACTGATCTGGAGCGCCCCCACTCACA	1133
XX	CC	various diseases including cancer e.g. adenocarcinoma and leukaemia,	QY	361	AlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr	380
XX	CC	immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease	Db	1134	GGTGGGTGCCACCTTACTTATCATGGCGGTGACGACGGGGACCATTTTACCATCACC	1193
XX	CC	and Parkinson's disease. The present sequence represents a polynucleotide	QY	381	ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla	400
XX	CC	probe of the invention. Note: The sequence data for this patent did not				
XX	CC	form part of the printed specification but was obtained in electronic				
XX	CC	format directly from USPTO at				
XX	CC	seqdata.uspto.gov/sequence.html?DocID=06500938B1				
XX	SQ	Sequence 3171 BP; 740 A; 903 C; 864 G; 664 T; 0 U; 0 Other;				

Alignment Scores:
 Pred. No.: 0 Length: 3171
 Score: 4369.00 Matches: 829
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Ds: 7 Gaps: 0

US-09-916-849A-1 (1-829) x ACA56750 (1-3171)

QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln 20
 Db 54 ATGGGCTCCCTCGTGGACCTCTCGCGTCTCTCCTCTCTCCAGGTTTGTGCTGCTGAC 113

Db 1194 ACCCACCTGAGACCAACCGGGCATCTGACACACCGAGGAGGGTTGGATTTTGAGGCC 1253
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAAACGACACACCTGTAGTTGAAGTGAACCAACGAGGCCCTTTTGTGCTGAAGCTC 1313
Qy 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCCACCATAGTGTCCAGTGGAGGATGTGAATGAGGACACCTGTGTTT 1373
Qy 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1374 GTCCACCCCTCCAAAGTCGTGTAGTCCAGAGGGCATCCCACTGGGAGGCTGTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTGTCAGAAAGACCCCTGACAGAGAGAAATCAAAAGATCAGCTACCGCATCTGAGA 1493
Qy 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCAGAGGTGGCTAGCCATGGACCCAGACAGTGGGAGGTACAGCTGTGGGACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGGAGTATGAGCATTTGTGAGGAACAACATCTATGAAGTCATGTGCTTGGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeuAsp 540
Db 1614 ATGGACAATGGAAGCCCTCCACACCTGCGACGCGGAACCCCTCTGCTAACACTGATGTAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGGCCAGTCCCTGAGCCCCGTGAGATCACCATCTGCAACCAAGCCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGGCCACGTGTGAACATCATCGGACAGGACCTGTCTCCACACACCTCCCTTTCCAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGluValAsp 600
Db 1794 GCCAGCTCAGATGACTCAGACATCTACTGGACGCGAGAGTCAACGAGGAGGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGTCTTGTCCCTGAAGAAGTTCCTGAAGCAGGATACATATGACGTGCACCTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTGACCATGGCAACAAAGAGCAGCTACCGTGTATCAGGCGCCACTGTGTCCGACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTyrTrpLysGlyGlyPheIleLeuProValLeu 660
Db 1974 CATGGCCATGTGGAACCTGCCCTGGACCTGGAAAGAGGTTTCATCTCCCTGTGTGTG 2033
Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuLeuValLeuLeuValArgLysLys 680
Db 2034 GGGCTGTCTGTGCTGTGCTGTCTCTCTCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 2093
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAGAGAGCCCT 2153
Qy 701 TyrGlyGluGlyGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCAAGAGGGGGTGGCAAGAGGACAGGACTATGACATCACCCAGCTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db 2214 GGTCTGAGGCGAGCGGAGGTGGTTCTCCGAATACGTGGACACCAACATCATCCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760

Db 2274 ACACCATGTACCGTCTAGGCCAGCCACCAACCCAGATGAATCGGCACTTTATAATTGAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGCGGCTAACACACAGACCCACGCCCCCTACGACACCCCTCTTGGTTC 2393
Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGGCAGCGGTCCGACCGCGTCCCTCGAGCTCCCTCACCTCCCTCCCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGACCGCTTCAAGAGCTGCA 2513
Qy 821 AspMetTyrGlyGlyGlyGlyGluAspAsp 829
Db 2514 GACATGTACGTGGCGGAGGAGCAGAC 2540
RESULT 7
ABZ24736
ID ABZ24736 standard; cDNA; 3171 BP.
XX
AC ABZ24736;
XX
DT 07-APR-2003 (first entry)
XX
DE Human P-cadherin cDNA.
XX
KW P-cadherin; human; colon cancer; colorectal cancer; cytostatic;
KW Gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 54..2543
FT /tag= a
FT /product= "Human P-cadherin"
XX
PN W0200297395-A2.
XX
PD 05-DEC-2002.
XX
PF 31-MAY-2002; 2002WO-US017109.
XX
PR 31-MAY-2001; 2001US-0294225P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Reinhard C, Klinger J, Jefferson AB, Escobedo J, Randazo F;
PI Winter J, Goodson R;
XX
DR WPI; 2003-140501/13.
DR P-PSDB; ABP58357.
XX
PT Inhibiting migration and proliferation of P-cadherin expressing cancer
PT for treating cancer, especially digestive cancer, characterized by
PT overexpression of P-cadherin, involves administering a P-cadherin
PT antagonist.
XX
PS Disclosure; Page 11-13; 129pp; English.
XX
CC The present sequence is the nucleic acid sequence for human P-cadherin
CC (placental cadherin), a calcium-dependent cellular adhesion protein. The
CC invention provides methods of treating or diagnosing cancers involving P-
CC cadherin expression using ligands that target P-cadherin, especially
CC human anti-P-cadherin antibodies. A claimed method of treating a cancer
CC characterised by the overexpression and/or upregulation of P-cadherin
CC comprises the administration of a P-cadherin antagonist, optionally
CC conjugated to a therapeutic agent. The migration, adhesion and/or
CC proliferation of the cancer is inhibited, and the method is especially
CC useful for treating or preventing a digestive cancer such as colon or
CC colorectal cancer. The antagonists may be an anti-P-cadherin antibody or
CC its fragment, a ribozyme or antisense oligonucleotide. A transgenic

Db 474 ATTTTCTACAGCATCAGGGGCGGGGCGAGACAGCCCTGAGGGTGTCTCGCTGTA 533
Qy 161 GluysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
Db 534 GAGAAGAGAGACAGCGCTGGTGTGGTGAATAAGCCACTGACCGGGAGAGATGGCAAG 593
Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 594 TATGAGCTCTTGGCCAGCTGTGTACAGAGATGGTGCCTCAGTGGAGGACCCCATGAAC 653
Qy 201 IleSerIleValThrAspGlnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 654 ATCTCCATCATCGTACCGACCAAGATGACCAAGCCCAAGTTTACCCAGGACCTTC 713
Qy 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 714 CGAGGAGTGTCTTAGAGGAGTCTTACCAAGTACTTCTGTGATGCGGTGACAGCCACA 773
Qy 241 AspGluAspAlaIleTyrThrTyrAsnGlyValValAlaIleTyrSerIleHisSerGln 260
Db 774 GATGAGGATGATGCATCTACACCTACATGGGGTGTCTTACTTCCATCCATAGCCAA 833
Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 834 GAACCAAGGACCCACACGACCTCATGTTACAAATCACCGAGCAGGACACCATCAGC 893
Qy 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db 894 GTCATCTCCAGTGGCTGGACCGGGAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
Qy 301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValGluIleLeuAsp 320
Db 954 ACAGACATGGATGGGAGCGGCTCCACACCGCAGTGGCAGTAGTGGAGATCCCTGAT 1013
Qy 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1014 GCCAATGACATGTCCCATGTTTGACCCCGAGAGTACGAGGCCCATGTGCTGAGAT 1073
Qy 341 AlaValGlyHisGluValAlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1074 GCAETGGGCCATGAGGTGCAGAGCTGACGGTCACTGATCTGGACGCCCCCACTACCA 1133
Qy 361 AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db 1134 GCGTGGCGTGCACCTACTTATCATGGCGGTGACGACGGGACCATTTTACCATCAC 1193
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1194 ACCCACCCTGAGAGCAACCGGCGATCTGACCAACCGAGGAGGTGGATTTGAGGCC 1253
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAACCGACACACCTGTGAGTGAAGTGACCAACGAGGCCCTTTTGTGCTGAAGCTC 1313
Qy 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACTCCACAGCACCATAGTGTGTCCACGTGGAGGATGTGAATGAGGCACCTGTGTT 1373
Qy 441 ValProSerLysValValGluValGlnGlyIleProThrGlyGluProValCys 460
Db 1374 GTCCCACTCCAAAGTGTGAGTCCAGAGGCGCATCCCACTGGGAGCGCTGTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACATGACAGACACCTTGACAGGAGATCAAAAGATGACGTACCGCATCTCAGA 1493
Qy 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCAGCAGGCTGCTAGCATGACCATGACACAGAGTGGGAGGTACAGCTGTGGGACCC 1553
Qy 501 LeuAspArgGluAspGluPheValArgAsnIleTyrGluValMetValLeuAla 520

Db 1554 CTCGACCCTGAGGATGAGCAGTGTGTGGAGAACAAATCTATGAAGTCATGGTCTTGGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db 1614 ATGGACATGGAAGCCCTCCACCACTGGCAGCGGAACCCCTTCTGTCTAAGACTGATGAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGGCCAGTCCCTGAGCCCGGTGAGATCAGATCAGATCTGCAACCAAGCCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGGCCACAGTGTGAAATCATCGACAGAGACCTGTCTCCCAACACCTCCCTTTCCAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 1794 GCGCAGCTCAGATGATCAGATCTACTGAGCGGAGGTCAACGAGGAAGGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGGTCTTGTCTCCTGAGAGTCTCTGAGCAGGATACATATGAGCTGCACCTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTGACCATGGCAACAAAGAGCAGCTGACGGTGTATCAGGGCCACTGTGTGAGCTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValLeu 660
Db 1974 CATGCCCATGTGAAACCTGCCCTGGACCCCTGGAAAGAGGTTCATCTCCCTGTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2034 GGGGCTCTCTGCTCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2093
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAAGGACCCCTCTACTCCAGAGATGACCCCGTGACACAGCTTTCTTCTAC 2153
Qy 701 TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGCGAAGAGGGGGTGGCGAAGAGACCCAGGACTATGACATCACCAGCTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db 2214 GGTCTGGAGGCCAGCGCGAGTGTCTCCGCAATGACGTGGCACCACCAACATCATCCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db 2274 ACACCATGTACTCTCTAGCGCCACCAACCCAGATGAATCGGCAACTTTAATTCAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGCGGTAAACACAGACCCACAGCCAGATGAATCGGCAACTTTAATTCAG 2393
Qy 781 AspTyrGluGlySerGlySerAspAlaLaserLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGCGGCGGTCCCGCGGTCCCTGAGCTCCCTCCTCCTCCTCCTCCTCCTCCTC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluIleThrGlySerArgPheLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCGCTTCAAGAAGCTGGA 2513
Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 2514 GACATGACGCTGGCGGAGGACGAC 2540

RESULT 9

ACH03801
ID ACH03801 standard; cDNA; 3777 BP.XX ACH03801;
XX DT 26-SBP-2003 (first entry)

Db 2056 GACCCAGCAGGCTGGCTAGCCATGACCCAGACAGTGGGAGGTACAGCTGTGGGACCC 2115
 Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
 Db 2116 CTCACCGTGAGGATGAGCAGTTGTGAGGACACACATCTATGAGTCATGCTTGGCC 2175
 Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuLeuLeuLeuLeu 540
 Db 2176 ATGACCAATGAGAGCCCTCCACACATGGCAGCGGAAACCCCTTCTGCTAACCATGATTGAT 2235
 Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
 Db 2236 GTCAACGACCATGGCCCATGCTCCCTGAGCCGCTGATCATCATCTGCAACCAAGCCCT 2295
 Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
 Db 2296 GTGGCCACAGTGTGAACATCACGGACAGGACCTGTCTCCACACCTCCCTTCCAG 2355
 Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
 Db 2356 GCCCAGCTCACAGATGACTCAGATCTTCTGAGCGCAGAGGTCAACGAGGAGGTGAC 2415
 Qy 601 ThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
 Db 2416 ACAGTGGTCTTGTCCCTGAGAGTCTCTGAAGCAGGATACATATGACGTGCACCTTCT 2475
 Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
 Db 2476 CTGTCTGACCATGGCAACAGACAGCTGACGGTGATCAGGGCCACTGTGTGCGACTGC 2535
 Qy 641 HisGlyHisValGluThrCysProGlyProTyrPheLeuLeuProValLeu 660
 Db 2536 CATGGCCATGTCGAACCTGCTGAGCCCTGGAGGGAGGTTTCATCTCCCTGTGCTG 2595
 Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuLeuValLeuLeuValArgLysLys 680
 Db 2596 GGGGCTGTCTCTGCT 2655
 Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
 Db 2656 CGAAGATCAAGAGCCCT 2715
 Qy 701 TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
 Db 2716 TATGSCGAAGAGGGGGTGGCGAAGAGGACGAGGACATGACATCACCAGCTCCACCGA 2775
 Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
 Db 2776 GGTCTGGAGGCGAGCGGAGGTGTCTCTCGCAATGACGTGGCACCACCAATCATCCCG 2835
 Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleGlu 760
 Db 2836 ACACCCATGTACCGTCTTAGCCAGCCACCCAGATGAATCGCACTTTAATATTGAG 2895
 Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
 Db 2896 AACCTGAAGGGCGGTAAACACAGACCCACACCCCGCCCTACGACACCCCTCTTGGTGTT 2955
 Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
 Db 2956 GACTATGAGGGCAGCGGTTCGAGCGCGCTCTGAGCTCCCTCCTCCTCCTCCTCCTCCT 3015
 Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLeuAla 820
 Db 3016 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCCGCTTCAAGAGCTGGCA 3075
 Qy 821 AspMetTyrGlyGlyGluAspAsp 829
 Db 3076 GACATGTACGGTGGCGGGAGGACGAC 3102

RESULT 10
 ABQ81547
 ID ABQ81547 standard; cDNA; 3205 BP.

XX ABQ81547;
 AC 30-DEC-2002 (first entry)
 XX
 DT
 XX
 DE Gene up-regulated in metastatic colorectal cancer.
 XX
 XX Colorectal cancer; metastasis; differential expression; cytostatic;
 KW diagnosis; gene therapy; vaccine; gene; ss.
 XX
 OS Homo sapiens.
 XX
 XX Location/Qualifiers
 FH 71..2560
 FT /cdg= a
 FT /product= "Metastatic colon cancer polypeptide"
 XX
 XX W0200268677-A2.
 XX
 XX 06-SEP-2002.
 XX
 XX 27-FEB-2002; 2002WO-US006001.
 XX
 XX 27-FEB-2001; 2001US-0272206P.
 PR 02-APR-2001; 2001US-0281149P.
 PR 17-APR-2001; 2001US-0284555P.
 XX
 XX (EOSB-) EOS BIOTECHNOLOGY INC.
 PA (OYCA-) UNIV CASE WESTERN RESERVE.
 XX
 XX Mack DH, Markowitz SD;
 XX
 XX WPI; 2002-698677/75.
 DR P-PSDB; ABP54683.
 XX
 XX New genes that are up- or down-regulated in colorectal cancer, useful for
 PT diagnosing colorectal cancer in a subject, or for identifying modulators
 PT of colorectal cancer-associated proteins and genes for treating
 PT colorectal cancer.
 XX
 XX Claim 5; Page 244-245; 260pp; English.
 XX
 XX The present sequence is the nucleotide sequence of a human gene that
 CC exhibits increased expression in metastatic colorectal cancer (MCC).
 CC samples. The gene is up-regulated in colon cancer-derived metastases
 CC compared to normal colon tissue. It is an example of claimed nucleic acid
 CC molecules that are up- or down-regulated in metastatic colorectal cancer
 CC cells. Such MCC-associated nucleic acids are useful in diagnostic and
 CC prognostic applications, in screening applications e.g. biochips, for
 CC identification of variant MCC-associated sequences, in informatics, for
 CC expression of MCC-associated proteins, in drug screening assays for
 CC identification of modulators of MCC-associated proteins, which are useful
 CC for treating this cancer, in gene therapy, as DNA vaccines, and as
 CC antisense or ribozyme modulators of MCC
 XX
 SQ Sequence 3205 BP; 751 A; 909 C; 875 G; 670 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 3205
 Score: 4361.00 Matches: 828
 Percent Similarity: 99.88% Conservative: 0
 Best Local Similarity: 99.88% Mismatches: 1
 Query Match: 99.82% Indels: 0
 DB: 6 Gaps: 0

US-09-916-849A-1 (1-829) x ABQ81547 (1-3205)

Qy 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuValCysTrpLeuGln 20

Db 71 ATGGGGCTCCCTCGTGGACCTCTCGGCTCTCCCTCTCCAGTTGTGGTGCAG 130

Qy 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40

Db 131 TCGCGCGCTCCGAGCGGTGCGGCGGTCTTCAGGGAGGCTGAAGTGACCTTGGAGGCG 190
Qy 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db 191 GAGGCGCGGAGCAGAGCGCGCGCGCGGTGGGAAAGTATTATCATGGGTGGCCTGGG 250
Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValAlaAsnGlyGluThr 80
Db 251 CAAGAGCGAGCTCTGTTAGCACTGATAATGATGACTTCACTGTGGCGAATGGCGAGACA 310
Qy 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerIleArg 100
Db 311 GTCCAGGAAAGAGGTCTACTGAAGAAAGGAATCCATTGAAGATCTTCCCATCCAAAGCT 370
Qy 101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
Db 371 ATCTTACGAAGACACACAGAGAGATTGGGTGGTTCCTCCAAATATCTGTCCTGAAAATGGC 430
Qy 121 LysGlyProPheProGluArgLeuAsnGlnLeuLysSerAsnLysAspArgAsnThrLys 140
Db 431 AAGGTCTCCCTTCCCCCAGAGACTGAATCAGCTCAAGTCTAATAAAGATAGAGACCAAG 490
Qy 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
Db 491 ATTTTCTACAGCATCACGGGCGCGGGGCGACAGCCCCCTGAGGGTGTCTTCGCTGTA 550
Qy 161 GluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
Db 551 GAGAAGGAGACAGAGCTGGTGTGTTGAATAAGCCACTGGACCGGAGGAGATTGCCAAG 610
Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 611 TATGAGCTCTTTGGCCAGCTGTGTCAAGAAATGGTGCCTCAGTGGAGAGCCCCATGAAC 670
Qy 201 IleSerIleIleValThrAspGluAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 671 ATCTCCATCATCGTGACCGCACAGATGACACACAGCCCCAAGTTTACCGAGGACACCTTC 730
Qy 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 731 CGAGGGAGTGTCTTAGAGGGAGTCTTACCAGTACTTCTGTGATGACAGGTGACAGCCACG 790
Qy 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
Db 791 GATGAGGATGATGCATCTACACCTACATGGGGTGGTGTGTTACTCATCCATAGCCAA 850
Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 851 GAACCAAGGACCCACACAGACCTCATGTTACCACTTACCCGGAGCACAGGCACCATCAGC 910
Qy 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db 911 GTCATCTCCAGTGGCCTGGACCGGAAAGTCCCTGAGTACACACTGACATCCAGGCC 970
Qy 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluIleLeuAsp 320
Db 971 ACACATGATGATGGGACGGCTCCACCACCGACGTGGCAGTGTGAGATCCTTGAT 1030
Qy 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1031 GCCAATGACAAATGCTCCATGTTTGACCCCCCAGAAAGTACAGGCCCATGTGCGCTGAGAT 1090
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1091 GCAGTGGGCCATGAGGTGCAGAGGCTGACGCTCACTGATCTGGACGCCGCCCACTCACCA 1150
Qy 361 AlaTrpArgAlaThrTyrLeuIleMetGlyLysAspGlyAspHisPheThrIleThr 380
Db 1151 GCCTGGCGTGCACCTTATCATGGCGGTGACGACGGGACCATTTTACCATCACC 1210
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1211 ACCACCTCGAGAGCAACACAGGCGATCTCTGACAAACAGGAAGGTTTGGATTTTGGAGGC 1270

Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1271 AAAAACCCAGCACACCCCTGTACGTGAAGTGACCAACAGAGCCCTTTTGTGCTGAAGCTC 1330
Qy 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
Db 1331 CCACCTCCACACGCCACCATAGTGGTCCAGCTGAGAGATGTGAATGAGGACCTGTGTTT 1390
Qy 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1391 GTCCACCCCTCCAAAGTCGTTGAGGTCCAGGAGGCGATCCCCACTGGGAGCGCTGTGTGT 1450
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1451 GTCTACATGTCAGAGAAGCCCTGCAAGGAGAAATCAAAAGATACGTACCCTCCCTGTGAGA 1510
Qy 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1511 GACCCAGCAGGCTGGCTAGCCATGGACCCAGACAGTGGGCGAGTCCACAGCTGTGGGCACC 1570
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1571 CTGACCGGTGAGGATGAGCAGTTTGTGAGGAACAACATCTATGAAGTCAATGCTTGGGC 1630
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db 1631 ATGGACAATGGAAGCCCTCCACCACTGGACCGGGAACCCCTTCGTACACACTGATTGAT 1690
Qy 541 ValAspAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1691 GTCAATGACCATGGCCCGCTGCTGAGCCCGCTGAGTACCATCACCATCTGCAACCAAGCCCT 1750
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1751 GTGGCGCAGGTGCTGACATCATCGGACAAAGACCTGTCTCCCAACACCTCCCTTTCAG 1810
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 1811 GCCAGCTCACAGATGACTCAGACATCTACTGGACGGCAGAGGTCAACGAGGAAGGTGAC 1870
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1871 ACAGTGTCTTGTCCCTGAAGAAGTCTCTGAAGCAGGATACATATGAGGTGACCTTTCT 1930
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1931 CTGTCTGCATGCAACAAACAGAGCAGCTGACGGTGTATCAGGGCCACTGTGTGCGACTGC 1990
Qy 641 HisGlyHisValGluThrCysProGlyProThrLysGlyGlyPheIleLeuProValLeu 660
Db 1991 CATGGCCATGTCCAAACCTGCGCCCTGGACCTTGAAGGGAGGTTTCATCTCCCTGTGCTG 2050
Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuValArgLysLys 680
Db 2051 GGGGCTGTCTGCTGTCTCTCTCTGCTGTGCTGTCTGCTTTGTTGTTGGTGAAGAAGAG 2110
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2111 CGGAAGATCAGAGAGCCCTCTCTACTCCAGAAAGATGACACCCGTGACAACTCTCTTAC 2170
Qy 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2171 TATGGCGAAGAGGGGGTGGCGAAGGACCGAGGACTATGACATCACCAGCTCCACCGA 2230
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db 2231 GGTCTGGAGGCGAGCGCGAGGTGCTTCTCCGCAATGACGTGGCACCACCAACCATCATCCG 2290
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db 2291 ACACCATGTACCGTCTCGGCCAGCCAAACCCAGATGAATCGGCAACTTTATATATTGAG 2350

Alignment Scores:	
Pred. No.:	0
Score:	4361.00
Percent Similarity:	99.88%
Best Local Similarity:	99.88%
Query Match:	99.82%
DB:	7
Length:	3205
Matches:	828
Conservative:	0
Mismatches:	1
Indels:	0
Gaps:	0

US-09-916-849A-1 (1-829) x ACC72821 (1-3205)

Qy	1	MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln	20
Db	71	ATGGGGCTCCCTGGTGGACCTCTCGGCTCTCTCTCTCCAGGTTTCTCGGTGCAG	130
Qy	21	CysAlaAlaSerGluProCysA-gAlaValPheArgGluAlaGluValThrLeuGluAla	40
Db	131	TGCGGCGCTCCGAGCGCTGCGGGCGGCTCTTCAGGGAGGCTGAAGTAGACCTTTGGAGCG	190
Qy	41	GlyGlyAlaGluGlnProGlyGlnAlaLeuGlyValPheMetGlyCysProGly	60
Db	191	GGAGGCGCGAGCAGGAGCCCGCGCCGCTGGGGAAAGATTTCATGGGCTGCCCTGGG	250
Qy	61	GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr	80
Db	251	CAAGAGCCAGCTCTGTTAGCACTGATATGATGACTTCACGTGCGGAATGGCGAGACA	310
Qy	81	ValGlnGluArgSerLeuLysGluArgAspProLeuLysIlePheProSerLysArg	100
Db	311	GTCCAGGAAGAAGGTCACCTCAAGGAAGGAATCCATTGAAGATCTTCCCATCCAAACGT	370
Qy	101	IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly	120
Db	371	ATCTTTAGCAAGACACAGAGAGATTTGGTGGTGTCTCCAAATATCTGCCCTGAANAATGGC	430
Qy	121	LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys	140
Db	431	AAGGGTCCCTTCCCCCAGAGACTGAATCAGCTCAAGTCTAATAAGATAGAGACACCAAG	490

141	IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGluGlyValPheAlaVal	Qy
141	IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGluGlyValPheAlaVal	Qy
491	ATTTTCTACAGCATCACGGGGCCGGGGCCAGACAGCCCCCTGAGGGTGTCCTCGCTGTA	Db
161	GluGlysglThrGlyTyrTrpLeuLeuLeuAsnLysProLeuAspArgGluGluLeuAlaLys	Qy
161	GluGlysglThrGlyTyrTrpLeuLeuLeuAsnLysProLeuAspArgGluGluLeuAlaLys	Qy

Db 551 GAGAAAGGACAGCGCTGGTTGTTTGTGTAATAAGCCACTGGACCGGGAGAGATTGCCAAG 611

Qy 181 TyrGluLeupheGlyHisalavalSerGluAsnGlyAlaSerValGluAspProMetAsn 200

Db	611	TATGAGTCTTTGGCCACGCTGTGTCAGAGAAATGGTGCCTCAGTGGAGGACCCCAAGAAC	870
Qy	201	IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe	220
Db	671	ATCTCCATCATCTGTGACCGACACAGATGACCAAGCCCAAGTTATCCAGGACACCTTC	730
Qy	221	ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr	240
Db	731	CGAGGAGTGCTTAGAGGGAGTGCTACAGGTACTTCTGTGATGCAGTGACAGCCACG	790
Qy	241	AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln	260
Db	791	GATGAGGATGATGCCATCTACACCTTACATGATGGGGTGTGCTTATCCATCATAGCCAA	850
Qy	261	GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer	280
Db	851	GAACCAAGGACCCACAGACTCATGTTTACCATTACCGGAGCACAGGCCACCATCAGC	910

Qy	21	CysAlaalaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla	40	Qy	361	ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgIysGlyLeuAspPheGluAla	400
Db	131	TGCCGGCCCTCCGAGCGCGTCCGGCGCGTCTTCAGGAGGCTGAGTGACCTTGGAGGGC	190	Db	1211	ACCCACCTTGAGAGCAACAGGGCATCTGTACAAACCCAGGAAGGGTTTTGGATTTTGAGGCC	1270
Qy	41	GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly	60	Qy	401	LysAsnGlnHisThrLeuThrValGluValThrAsnGluAlaProPheValLeuLysLeu	420
Db	191	GGAGCGCGGAGCAGGAGCGCGCCAGCGCTGGGGAAGTATTCATGGGCTGCCCTGGG	250	Db	1271	AAAAACCAAGCAACCTGTACCTTGAAGTGACCAACGAGGCCCCCTTTTGTGCTGAASCTC	1330
Qy	61	GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr	80	Qy	421	ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe	440
Db	251	CAAGAGCCAGCTCTGTTTAGCAGTGAATGATGACTTCATGTGCGGAATGGCAGACA	310	Db	1331	CCAACCTCCACAGCCACCATAGTGTCCACGTGAGGATGTGAATGAGGCACCTGTGT	1390
Qy	81	ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg	100	Qy	441	ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys	460
Db	311	GTCCAGGAAAGAAGGTCACTGAAGGAAGGAATCCATTGAAGATCTTCCCATCAAAAGT	370	Db	1391	GTCCACACCTCCAAAGTCGTTGAGGTCAGGAGGCGATCCCACTGGGAGCCCTGTGT	1450
Qy	101	IleLeuArgArgHisLysArgAspTyrValValAlaProIleSerValProGluAsnGly	120	Qy	461	ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg	480
Db	371	ATCTTACGAGACACAAGAGATGGTGGTGTCTCCAAATATCTGTCCCTGAAATGGC	430	Db	1451	GTCTACACTGCAGAGACCTTGACAGGAGAAATCAAAAGATCAGCTACCGCATCTCGAGA	1510
Qy	121	LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys	140	Qy	481	AspProAlaGlyTyrPheAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr	500
Db	431	AAGGTCCTCTTCCCCAGAGACTGAATCAGCTCAAGTCTAATTAAGATAGACACCAAG	490	Db	1511	GACCCAGCAGGCTGGCTAGCCATGACCCACAGTGGGCGAGGTACAGCTGTGGGACC	1570
Qy	141	IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal	160	Qy	501	LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla	520
Db	491	ATTTTCTACAGCATCACGGGCGCGGGGCGACAGACGCCCTGTAGGGGTGTCTCGCTGTA	550	Db	1571	CTCGACCTGTAGGATGAGCAGTTTGTGAGGAACAACATCTATGAAGTCATGGTCTTGCC	1630
Qy	161	GluLysGluThrGlyTyrPheLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys	180	Qy	521	MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeuAsp	540
Db	551	GAGAGGAGACAGGCTGCTGTGTGAATAAGCCATCGACCCGGGAGAGATGGCCAG	610	Db	1631	ATGGCAATGAAAGCCCTCCACCACTGGCAGCGGAACCTCTCTGCTAACACTGATTGAT	1690
Qy	181	TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn	200	Qy	541	ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro	560
Db	611	TATGAGCTCTTTGGCCACGCTGTGTACAGAGATGGTGCCTCAGTGGAGGACCCCATGAAC	670	Db	1691	GTCAATGACCATGGCCCACTCCCTGAGCCCGGTGAGATCACCATCTGCAACCAAGCCCT	1750
Qy	201	IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe	220	Qy	561	ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln	580
Db	671	ATCTCCATCATCGTACCGACAGAGTACACCAAGGCCCAAGTTTACCAGGACACCTTC	730	Db	1751	GTGGCCAGGTCTGAACATCAGCGACAAGGACCTGTCTCCCCACACCTCCCTTTCCAG	1810
Qy	221	ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr	240	Qy	581	AlaGlnLeuThrAspAspSerAspIleTyrTyrThrAlaGluValAsnGluGluGlyAsp	600
Db	731	CGAGGAGTGTCTAGAGGAGTCTTACAGGACTTCTGTGTGATGACAGGTGACGCCACG	790	Db	1811	GCCCACTCACAGATGACTCAGACATCTACTGACCGCGAGAGGTCAACGAGGAGGTGAC	1870
Qy	241	AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln	260	Qy	601	ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer	620
Db	791	GATGAGGATGATGCCATCTACACCTTACAAATGGGGTGGTGTCTTACTCATCCATAGCCAA	850	Db	1871	ACAGTGTCTTGTCCCTGAAGAAGTTCTGTAAGCAGGATACATATGACGTGACCTTTCT	1930
Qy	261	GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer	280	Qy	621	LeuSerAspHisGlyAsnLysGlnGlnLeuThrValIleArgAlaThrValCysAspCys	640
Db	851	GAACCAAGGACCCACAGACTCATGTTTACCATTACCCGAGACAGGACCATCAGC	910	Db	1931	CTGTCTGACCATGGCAACAAGAGCAGCTGACGGTGATCAGGGCCACCTGTGTGCCACTGC	1990
Qy	281	ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGluAla	300	Qy	641	HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu	660
Db	911	GTCACTCCAGTGGCTTGACCGGGAAAGTCCCTGAGTACACACTGACCATCCAGGCC	970	Db	1991	CATGGCCATGTGAACCTGCCCTGACCCCTGAAGGAGGAGGTTTCATCTCCTCCTGTGT	2050
Qy	301	ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValGluIleLeuAsp	320	Qy	661	GlyAlaValLeuAlaLeuLeuPheLeuLeuLeuValLeuLeuLeuValArgLysLys	680
Db	971	ACAGACATGATGGGAGCGGCTCCACCAACCGGAGTGGGAGTAGTGAGATCTTGAT	1030	Db	2051	GGGGCTGTCTGGCTGTCTGTCTCTCTGCTGGTGGTGGTCTTTTGTGTGTGAAGAAGAG	2110
Qy	321	AlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn	340	Qy	681	ArgLysValIleLysGluProLeuLeuLeuProGluAspThrArgAspAsnValPheTyr	700
Db	1031	GCCATGACAAATGCTCCCATGTTTGAACCCCAAGAGTACGAGGCCATGTGCTGAGAAT	1090	Db	2111	CGGAAGATCAGAGGCCCTCTCTACTCCCAAGAAGTACACCCGTCGACACCTCTTCTAC	2170
Qy	341	AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro	360	Qy	701	TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg	720
Db	1091	GCAGTGGGCGCATGAGGTGACGGTCACTGATCTGGAGCGCCCACTCACC	1150	Db	2171	TATGGCAAGAGGGGGTGGCAAGAGGACCAAGGACTATGACATCACCCAGTCCACCGA	2230
Qy	361	AlaTyrArgAlaThrTyrIleIleMetGlyGlyAspAspGlyAspHisPheThrIleThr	380	Qy	721	GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro	740
Db	1151	GGGTGGCGTGGCCACCTTATCATGGCGGTGACGAGGGACCACTTTTACCATCACC	1210	Db	2231	GGTCTGGAGGCCAGGCGGAGGTGTTCTTCGCAATGAGCTGGCACCACCACTATCCCG	2290
Qy				Qy	741	ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu	760

Db	2291	ACACCCATGTACCGTCTCTCGGCCAGCCAAACCCAGATGAAATCGGCACACTTTATAATTGAG	2350
Qy	761	AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe	780
Db	2351	AACCTGAAGCGGGCTAACACAGACCCACACAGCCCGCCTACGACACCCCTCTGGTCTTC	2410
Qy	781	AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer	800
Db	2411	GACTATGAGGCACGGCTCCGACCGCGCGTCCCTGAGCTCCCTCACCTCCTCGGCTCC	2470
Qy	801	AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLysLeuAla	820
Db	2471	GACCAAGACCAGATTACGATTATCTGACGAGTGGGCGCGCTTCAGAGCTGGCA	2530
Qy	821	AspMetTyrGlyGlyGlyGluAspAsp	829
Db	2531	GACATGTACGGTGGCGGGAGGACGAC	2557
RESULT 14			
AAD27637			
ID	AAD27637 standard; DNA; 3219 BP.		
XX	AAD27637;		
XX	18-APR-2002 (first entry)		
DT	Human colon specific gene #17.		
XX	Human; colon specific gene; CSG; imaging; colon cancer; metastasis;		
KW	screening; chromosomal identification; transgenic animal; cytostatic;		
KW	gene therapy; p cadherin; ds.		
XX	Homo sapiens.		
OS	WO200200939-A2.		
XX	03-JAN-2002.		
XX	28-JUN-2001; 2001WO-US020724.		
XX	28-JUN-2000; 2000US-0214515P.		
PR	(DIAD-) DIADEXUS INC.		
XX	Macina RA, Pillai R;		
XX	WPI; 2002-139938/18.		
DR	Detecting, diagnosing, monitoring, staging, prognosticating, imaging and		
XX	treating colon cancer using colon-specific genes.		
XX	Claim 1; Page 127-128; 135pp; English.		
XX	The invention relates to colon specific gene (CSG) and their polypeptides		
CC	used in a method of diagnosing, monitoring, staging, imaging and treating		
CC	colon cancer. CSG is useful for diagnosing the presence and metastasis of		
CC	colon cancer. It can be used for staging colon cancer in a patient having		
CC	colon cancer where an increase in determined CSG levels in the patient		
CC	compared to normal human control is associated with a cancer which is		
CC	progressing and a decrease in CSG levels is associated with a cancer		
CC	which is regressing or in remission. CSG is useful for identifying		
CC	potential therapeutic agents for use in imaging and treating colon cancer		
CC	by screening compounds that binds to, or decrease expression of CSG		
CC	relative CSG in the absence of the compound. CSG is used to identify		
CC	subjects having or at risk of developing a disease or disorder associated		
CC	with increased levels of CSG. CSG sequences are also used for chromosomal		
CC	identification. CSG sequences can also be used for producing non-human		
CC	transgenic animals. The methods can also be used to detect genetic		
CC	lesions or mutations in CSG. The identification of CSG is also useful in		
CC	the rational design of the new therapeutics for imaging and treating		
CC	cancers. CSG is used in gene therapy. The present sequence is human colon		
CC	specific gene encoding p cadherin		


```
QY 578 oPheGlnAlaGlnLeuThrAspSerAspIleTyrTyrThrAlaGluValAsnGluG1 598
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1953 CTTACAGCAAGACTTACACACAGCGGCAAGTGTCAACTGGACCATCGAGTACATGACCC 2012
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 598 uGly---AspThrValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspVa 617
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2013 AGCTCGTGAATCTCTAATTTGAAGCCAAAGAAAACCTTGAAGTTGGTGACTACAAAAT 2072
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 617 lHisLeuSerLeuSerHisGlyAsnLysGluLeuThrValIleArgAlaThrVa 637
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2073 AAATCTCAAGCTCACAGATAACCAAGAACAGAGTGCACCCCTATATGTGTTGT 2132
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 637 lCysAspCysHisGlyHisValGluThrCys-----ProGlyProTyr---LysGlyG1 654
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2133 GTGGAGCTGCGAAGGTGTCGTCACAGCTGCAAGAGGACGGCGCTTACGCCGAAGCAGG 2192
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 654 yPhe-----IleLeuProValLeuGlyAlaValLeuAlaLeuPheLeuLe 670
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2193 CTTGAGGTTCTCGCATCTTGGGCATCTCGAGGAATCCTCGCTCTACTAATCCTGAT 2252
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 670 uLeuValLeuLeuLeuValArgLysLysArgLysIleLysGluProLeuLeuLeuPr 690
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2253 TCTGCTGCTTCTGCTATTGTTGGAGGAGAGGGTGTCAAGAGCCCTTACTTCCCCC 2312
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 690 oGluAspAspThrArgAspAsnValPheTyrTyrGlyGluGluGlyGlyGluGluAs 710
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2313 AGAGATGACACCCCGGCAATGTTTATTACTATGATGAAGAGGAGGTGGAGAGGAGGA 2372
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 710 pGluAspTyrAspIleThrGlnLeuHisArgGlyLeuGluAlaArgProGluValValLe 730
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2373 TCAGGACTTTCGACTTGACCCAGTTGACAGGGGGCTGGATGCTCGGCTGNAAGTACT-- 2430
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 730 uArgAsnAspValAlaProThrIleIleProThrProMetTyrArgProArgProAlaAs 750
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2431 -CGCAATGATGTGGCCCAACCTCTCGAGTGTGGCCAGTATCGGCCCGCCCTGCCAA 2489
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 750 nProAspGluIleGlyAsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAspProth 770
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2490 TCCTGATGAATGTGAAACTTTATTGATGAACCTGAAGGCAGCGGACTGACCCCTAC 2549
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 770 rAlaProProTyrAspThrLeuLeuValPheAspTyrGluGlySerGlySerAspAlaAl 790
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2550 TGCCTCCTCTTATGACTCTCTGCTGCTGTTGACTATGAAGGAAGCGGTTCTGAAGCTGC 2609
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 790 aSerLeuSerSerLeuThrSerSerAlaSerAspGlnAspGlnAspTyrAspTyrLeuAs 810
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2610 TAGTCTGAGCTCCTTGAACTCCTCAGAGTCCAGACCAAGACCAGGACTATGACTACCTGAA 2669
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 810 nGluTyrGlySerArgPheLysLysLeuAlaAspMetTyrGlyGlyGlyGluAspAsp 829
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2670 TGAATGGGGCAATCGCTTCAAGAGAGTGGCGGACATGTATGGAGGTGGCGAGGACGAC 2727
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: September 22, 2004, 21:44:54
Job time : 864 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 22, 2004, 21:23:38 ; Search time 142 Seconds
(without alignments)
3239.819 Million cell updates/sec

Title: US-09-916-849a-1

Perfect score: 4369
Sequence: 1 MGLPRGLASLLQLVCWLQ.....NEWGSRFKKLADMYGGEDD 829

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_p2n.model -DEV=xlh
-C=/cgn2_1/USPTO_spool_h/US0916849/runat_21092004_145804_7033/app_query.fasta_1.967
-DB=Issued Patents NA -QMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0916849@cgn 1 69 @runat_21092004_145804_7033 -NCPUS=6 -ICFUS=3
-DEV -MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-NO_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: Issued Patents NA:*
- 2: /cgn2_6/ptdata/2/ina/5A.COMB.seq:*
- 3: /cgn2_6/ptdata/2/ina/5B.COMB.seq:*
- 4: /cgn2_6/ptdata/2/ina/6A.COMB.seq:*
- 5: /cgn2_6/ptdata/2/ina/6B.COMB.seq:*
- 6: /cgn2_6/ptdata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4369	100.0	3171	4	US-09-016-434-1348
2	2349.5	53.8	2808	1	US-08-237-919-1
3	2349.5	53.8	2808	4	US-08-732-429-1
4	2349.5	53.8	2808	4	US-09-798-267-1
5	2349.5	53.8	2808	5	PCT-US95-05518-1
6	1621.5	37.1	3048	1	US-08-188-228-47
7	1621.5	37.1	3048	1	US-08-332-643-41
8	1621.5	37.1	3048	1	US-08-332-638-47
9	1025.5	23.5	2690	1	US-08-188-228-61
10	1025.5	23.5	2690	1	US-08-332-643-55
11	1025.5	23.5	2690	1	US-08-332-638-61
12	977.5	22.4	2779	1	US-08-474-067-3

13	977.5	22.4	2779	2	US-08-474-068A-3
14	977.5	22.4	2779	2	US-08-472-481-3
15	976.5	22.4	2880	4	US-09-016-434-1125
16	976	22.3	3959	1	US-08-474-067-1
17	976	22.3	3959	2	US-08-474-068A-1
18	976	22.3	3959	2	US-08-472-481-1
19	968.5	22.2	3712	2	US-08-738-349-3
20	968	22.2	4125	4	US-09-620-312D-270
21	966.5	22.1	2625	1	US-08-188-228-57
22	966.5	22.1	2625	1	US-08-332-643-51
23	966.5	22.1	2625	1	US-08-332-638-57
24	964.5	22.1	3552	4	US-09-643-597-126
25	964.5	22.1	3552	4	US-09-480-884A-126
26	964.5	22.1	3552	4	US-09-542-615A-126
27	964.5	22.1	3552	4	US-09-606-421B-126
28	964.5	22.1	3552	4	US-09-221-107-126
29	951.5	21.8	3581	2	US-08-738-349-1
30	933	21.4	3136	1	US-08-188-228-41
31	933	21.4	3136	1	US-08-332-638-41
32	926.5	21.2	2550	1	US-08-188-228-53
33	926.5	21.2	2550	1	US-08-332-643-47
34	926.5	21.2	2550	1	US-08-332-638-53
35	924	21.1	3514	2	US-08-738-349-5
36	914.5	20.9	2521	1	US-08-188-228-59
37	914.5	20.9	2521	1	US-08-332-643-53
38	914.5	20.9	2521	1	US-08-332-638-59
39	853.5	19.5	2490	1	US-08-188-228-45
40	853.5	19.5	2490	1	US-08-332-638-45
41	786	18.2	4000	4	US-09-016-434-1434
42	786	18.2	4000	4	US-09-023-655-1452
43	794	18.2	3164	1	US-08-188-228-49
44	794	18.2	3164	1	US-08-332-643-43
45	794	18.2	3164	1	US-08-332-638-49

ALIGNMENTS

RESULT 1

US-09-016-434-1348
Sequence 1348, Application US/09016434
Patent No. 6500938

GENERAL INFORMATION:

APPLICANT: Janice Au-Young

APPLICANT: Jeffrey J. Seilhamer

TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING

TITLE OF INVENTION: PATHWAY GENE EXPRESSION

NUMBER OF SEQUENCES: 1490

CORRESPONDENCE ADDRESS:

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 PORTER DRIVE

CITY: PALO ALTO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/016,434

FILING DATE: HEREWITH

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Zeller, Karen J

REGISTRATION NUMBER: 37,071

REFERENCE/DOCKET NUMBER: PA-0002 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (650) 855-0555

QY 98 rlysaargileLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGI 118
DB 531 CTCTCCTGGCCTCAGAACAGACAGAGAGACTGGGTATTCTCTCCATCAGCTGCCAGA 590
QY 118 uAsnGlyLysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAs 138
DB 591 AAATGAAAGAGCCCAATTCCTTAAACCTGGTTCAGATCAATCCACAAAGACAAAGA 650
QY 138 pThrLysIlePheThrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPh 158
DB 651 AGGCAAGGTTTCTACACATCATCTGCGCAAGAGAGCTGACACCCCTGTGTGTCTT 710
QY 158 eAlaValGluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluI 178
DB 711 TATTATTGAAGAGAAACAGAGTGGCTGAAGTGCAGAGCTCTGGATAGAGAACGCAT 770
QY 178 eAlaLysThrGluLeuPheGlyHisAlaValSerGluAsnGlnGlyAlaSerValGluAspPr 198
DB 771 TGCCACATACATCTCTCTCTCAGCTGTGTCTCAACCGGAATGCAGTTGAGATCC 830
QY 198 oMetAsnIleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAs 218
DB 831 AATGGAGATTTGTATCAGGTACCGATCAGATGACAAAGCCCGAATTCACCCAGGA 890
QY 218 pThrPheArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValTh 238
DB 891 GGTCTTTAAGGGTCTGTATGAGAGGTGCTCTTCCAGGAACCTCTGTGTAGAGGTAC 950
QY 238 rAlaThrAspGluAspAlaIleThrThrTyrAsnGlyValValAlaIleThrSerIleHi 258
DB 951 AGCCACAGCCGACGATGATGTAACCATCAATGCCGCCATCGCTTACACCATCT 1010
QY 258 sSerGlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyTh 278
DB 1011 CAGCCACAGATCCTGAGCTCCCTGACAAAATATGTTCCACCATTAACAGGAACAGGAT 1070
QY 278 rIleSerValIleSerSerGlyLeuAspArgGluLysValProGluThrThrLeuThrI 298
DB 1071 CATCAGTGTGTCCACCTGAGCTGGAGCGGAGAGATTTCCCTACGTATACCTCTGTGTG 1130
QY 298 eGlnAlaThrAspMetAspGlyAspGlySerThrThrAlaValAlaValGluI 318
DB 1131 TCACCTGCTGACCTTCAGGTGAGGGGTAAAGCACACACACACACACTGTGTACACAGT 1190
QY 318 eLeuAspAlaAsnAspAlaProMetPheAspProGlnLysThrGluAlaHisValPr 338
DB 1191 CACTGACACCAACCAATATCTCCGATCTTCAATCCACACACAGGTACAAAGGTGCTGCC 1250
QY 338 oGluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAs 358
DB 1251 TGAAACAGAGCTAACGCTGTAATACCACTGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1310
QY 358 nSerProAlaTrpArgAlaThrThrLeuIleMetGlyLysAspGlyAspHisPheTh 378
DB 1311 TACCCACAGCTGGAGGCTGATACACCATTTG---AATGATGATGCTGCGACAAATTTG 1367
QY 378 rIleThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPh 398
DB 1368 CGTCACCAACAATCAGTGAACACAGATGGATTTTGAACAAACGAAAGGGCTGGATTT 1427
QY 398 eGluAlaLysAsnGlnHisThrLeuThrValGluValThrAsnGluAlaProPheValLe 418
DB 1428 TGAGCCCAAGCAGCAGTACATCTTACACGTAGCAGTACACGAATGTGTGTACCTTTGAGGT 1487
QY 418 uLysLeuProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaPr 438
DB 1488 CTCCTCTCACCACCTCCACAGCCCGCTCACCGTGGATGTGTGGATGTGAATGAAGGCC 1547
QY 438 oValPheValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluPr 458
DB 1548 CATCTTTGTGCTCTCTGAAAGAGAGTGGAGTGTCCGAGGACTTTTGGCGTGGCCAGGA 1607

QY 458 oValCysValThrThrAlaGluAspProAspLys---GluAsnGlnLysIleSerTyrAr 477
DB 1608 AATCATCATCTACATGCCCCAGAGGACAGACATTTATGGAAACAGAAAATACATATCG 1667
QY 477 gIleLeuArgAspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAl 497
DB 1668 GATTTGGAGAGACACTCGCAACTGGCTGGAGATTAAATCGGACACTGGTGCCATTCCAC 1727
QY 497 aValGlyThrLeuAspArgGluAspGluGlnPheValArgAsnAsnIleThrGluValMe 517
DB 1728 TCGGGCTGAGCTGGACAGGGAGATTTTGACGACGTGAAGAACACGCTACACAGCCCT 1787
QY 517 tValLeuAlaMetAspAsnGlySerProProThrThrThrGlyThrGlyThrLeuLeuTh 537
DB 1788 AATCATAGCTACAGACAATGGTTCCTCAGTGTCTACTGGACAGGACACTTCTCTGAT 1847
QY 537 rIleuLeAspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAs 557
DB 1848 CTTGTCTGTATGTGAATGACACACGCCCCCATACAGAACCTCGAATATTTCTTCTGTGA 1907
QY 557 nGlnSerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSe 577
DB 1908 GAGGATCCAAAGCCTCAGTTCATATAACATTCATGATGACAGACTTCTCCCAATACATC 1967
QY 577 rProPheGlnAlaGlnLeuThrAspSerAspIleTyrTrpThrAlaGluValAsnGl 597
DB 1968 TCCTCTTACAGCAGAACTAAACACACGGCGAGTGCCTCAACTGGACCATTCAGTACACGA 2027
QY 597 u---GluGlyAspThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAs 616
DB 2028 CCAACCCAAAGATCTATCATTTTGAAGCAAGATGCGCTTAGAGTGGGTGACTACAA 2087
QY 616 pValHisLeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaTh 636
DB 2088 AATCAATCTCAAGCTCATGATTAACCAAGATAAAGACCAAGTGACCATTTAGAGTCA 2147
QY 636 rValCysAspCysHisGlyHisValGluThrCys-----ProGlyProTrpLysGlyGl 654
DB 2148 CTTGTGTGACTGTGAAGGGGCGCGCGGTCTGTAGGAAGGCACAGCTCTGCAAGCAGG 2207
QY 654 yPhe-----IleLeuProValLeuGluValAlaValLeuAlaLeuLeuPheLeuLe 670
DB 2208 ATTGCAAATTCCTGCCATCTCTGGGATTTCTGGAGGAATTTCTTGTCTTGAATTCAT 2267
QY 670 uLeuValLeuLeuLeuValArgLysLysArgLysLysLysGluProLeuLeuLeuPr 690
DB 2268 TCTGTCTGTCTGTCTTCTTCTCGAGAGAGCGGTGCTCAAGAGAGCGCTTACTGCCCCC 2327
QY 690 oGluAspAspThrArgAspAsnValPheTyrTyrGlyGluGlyGlyGlyGluGluAs 710
DB 2328 AGAGATGACACCCGGGACACAGTTTATCTATGATGAAGAGAGCGGCGGAGAGAGGA 2387
QY 710 pGlnAspTrpAspIleThrGlnLeuHisArgGlyLeuGluAlaArgProGluValValLe 730
DB 2388 CCAGGACTTTCATTTGAGCCAGCTGCACAGGGCGCTGACGCTCGGCTGAAGTACT-- 2445
QY 730 uArgAsnAspValAlaProThrIleProThrProMetTyrArgProArgProAlaAs 750
DB 2446 -CGTAACAGAGTTTGCACCAACCTCATGAGTGTCCCGGATCTTCTCCCGCCCTGCCA 2504
QY 750 nProAspGluIleGlyAsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAspProTh 770
DB 2505 TCCCGATGAAATTTGAAATTTTATTTGATGAAATCTGAAAGCGGCTGATACTGACCCAC 2564
QY 770 rAlaProProThrAspThrLeuLeuValPheAspTyrCluGlySerGlySerAspAlaAl 790
DB 2565 AGCCCCGCTTATGATTTCTGTCTGTGTGTGACTATGAAGAGAGCGGCTTCCGAGCTGC 2624
QY 790 aSerLeuSerSerLeuThrSerSerAlaSerAspGlnAspGlnAspTyrAspTyrLeuAs 810
DB 2625 TAGTCTGAGCTCCCTGAACTCCTCAGAGTGCAGAAAGACAGGACTATGACTACTTGA 2684
QY 810 nGluTrpGlySerArgPheLeuLysLysLeuAlaAspMetTyrGlyGlyGlyGluAsp 828

QY 298 eGlnAlaThrAspMetAspGlyAspGlySerThrThrAlaValAlaValGluI 318
DB 1131 TCAAGCTGCTGACCTTCAAGGTGAGGGGTTAAGCACACACACAGCTGTGATCAGAT 1190
QY 318 eLeuAspAlaAsnAspAlaProMetPheAspProGlnIlySerThrGluAlaHisValPr 338
DB 1191 CACTGACACCAACGATAATCTCGATCTTCAATCCACACGACGACGAGGTGAGGTGCC 1250
QY 338 oGluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAs 358
DB 1251 TGAGAACGAGGCTACGTCGTAATCAACACACGACGACGATGCTGATGCTGATGCTGAT 1310
QY 358 nSerProAlaThrArgAlaThrThrLeuIleMetGlyGlyAspAspGlyAspHisPheTh 378
DB 1311 TACCCACGCGTGGGAGGCTGTATACACCATATG---AATGATGATGCTGACAAATTTGT 1367
QY 378 rIleThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArgIlySerGlyLeuAspPh 398
DB 1368 CGTCACCAACATCAAGTGAACACGATGCGATTTTGAACACACAAAGGCTTGATTT 1427
QY 398 eGluAlaIlyAsnGlnHisThrLeuThrValGluValThrAsnGluAlaProPheValle 418
DB 1428 TGAGGCCAAGCAGCAGTACATTTACACGTACGATGACGATGATGATGATGATGATGAT 1487
QY 418 uIlySerProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaPr 438
DB 1488 CTCCTCTACACCTCCACAGCCCGTCCGCTGGATGCTGATGATGATGATGATGATGATGAT 1547
QY 438 oValPheValProProSerLysValValGluValGlnGlyIleProThrGlyGluPr 458
DB 1548 CATCTTTGCTGCTCTGAAAGAGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 1607
QY 458 oValCysValThrAlaGluAspProAsPlys---GluAsnGlnIlySerThrAr 477
DB 1608 AATCATCTCTACATCTGCCAGGAGCCAGACATTTATGAAACAGAAATATATATATG 1667
QY 477 gIleLeuArgAspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAl 497
DB 1668 GATTGGAGAGACACTCCCACTGCTGGATTAATCCGACACACTGCTGCTGCTGCTGCTGCT 1727
QY 497 aValGlyThrLeuAspArgGluAspGluGlnPheValArgAsnIleThrGluValme 517
DB 1728 TCGGCTGAGCTGGACAGGAGGATTTGAGCAGCTGGAAGAACACGACGATACAGCCCT 1787
QY 517 tValLeuAlaMetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuTh 537
DB 1788 AATCATAGCTACAGACATGTTCTCCAGTTGCTACTGGAACAGGACACTTCTGCTGAT 1847
QY 537 rLeuIleAspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAs 557
DB 1848 CCGTCTGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1907
QY 557 nGlnSerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSe 577
DB 1908 GAGGAATCAAGCTCAGGTATTAATATTCATGATGACGACCTTCTCCCATATATC 1967
QY 577 rProPheGlnAlaGlnLeuThrAspAspSerAspIleThrTrpThrAlaGluValAsnGl 597
DB 1968 TCCCTTCACAGCAGAACTAAACACACGCGGAGTGCCTCACTGACCACTTCACTCAACAGA 2027
QY 597 u---GluGlyAspThrValValLeuSerLeuIlySerPheLeuLysGlnAspThrTrAs 616
DB 2028 CCCAACCCCAAGAAATCTATCTTTTGAACCAAGATGCGCTTAGAGGTGGTGTACTACAA 2087
QY 616 pValHisLeuSerLeuSerAspHisGlyAsnLysGlnLeuThrValIleArgAlaTh 636
DB 2088 AATCAATCTCAAGCTCATGATAACCAAGATAAAGCAACCAAGTACCACCTTAGAGGTGAG 2147
QY 636 rValCysAspCysHisGlyHisValGluThrCys-----ProGlyProThrLysGlyGl 654
DB 2148 CGTGTGTGACTGTGAAGGGCGCGCGCTGTGTAGGAAGCAGCAGCTGTGGAAGCAGG 2207

QY 654 yPhe-----IleLeuProValIleGlyAlaValLeuAlaLeuPheLeuLe 670
DB 2208 ATTGCAAAATCTGTCATCTGGGATTTCTGGAGGAATTTCTGTTTGTCTAATCTGAT 2267
QY 670 uLeuValLeuLeuLeuValArgLysIlySerGlyIleLysGluProLeuLeuLeuPr 690
DB 2268 TCTGCTGCTCTGCTGTTTCTCGGAGGAGAGCGGTGTCAAGAGAGCCTTACTTGCCTCC 2327
QY 690 oGluAspAspThrArgAspAsnValPheThrThrGlyGluGluGlyGlyGlyGluAs 710
DB 2328 AGAGGATGACACCGGACACCGTTTATCTATCTATGATGAAGAGGAGCGGAGAGAGA 2387
QY 710 pGlnAspTyrAspIleThrGlnLeuHisArgGlyLeuGluAlaArgProGluValValle 730
DB 2388 CAGAGCTTTGATCTGAGCCAGCTGCACAGGGCTCGACGCTCGCTCGCTGAAGTACT-- 2445
QY 730 uArgAsnAspValAlaProThrIleIleProThrProMetTyrArgProArgProAlaAs 750
DB 2446 -CGTAACGACGTTGACCAACCTCATGAGTGTCCCGCGTATCTTCCCGCTGCGCAA 2504
QY 750 nProAspGluIleGlyAsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAspProTh 770
DB 2505 TCCGATGAATTTGAATTTTATGATGAATAATCTGAAGCGGCTGATCTGACCTGAC 2564
QY 770 rAlaProProTyrAspThrLeuLeuValPheAspTyrGluGlySerGlySerAspAlaAl 790
DB 2565 AGCCCGCTTATGATTTCTCTGCTGTTGACTATGAAAGAGCGGTCCGAGCTGC 2624
QY 790 aSerIleuSerSerLeuThrSerSerAlaSerAspGlnAspGlnAspTyrAspTyrLeuAs 810
DB 2625 TAGTCTGAGCTCCCTGAACTCTCAGAGTCAGACAAAGACCAAGGACTATGACTACTTGA 2684
QY 810 nGluTrpGlySerArgPheLysLeuAlaAspMetTyrGlyGlyGlyGluAsp 828
DB 2685 CGAATGGGGCAATCGCTTCAAGAGCTGGCTGACATGTACGAGGCGCGAGGAC 2739
RESULT 4
US-09-798-267-1
; Sequence 1, Application US/09798267
; Patent No. 6406870
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael
; APPLICANT: Ceppek, Karyn
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin
; FILE REFERENCE: L0560/7008RP
; CURRENT APPLICATION NUMBER: US/09/798,267
; CURRENT FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 08/237,919
; PRIOR FILING DATE: 1994-05-03
; PRIOR APPLICATION NUMBER: PCT/US 95/05518
; PRIOR FILING DATE: 1995-05-03
; PRIOR APPLICATION NUMBER: US 08/732,429
; PRIOR FILING DATE: 1996-11-01
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2808
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(2742)
; NAME/KEY: mat_peptide
; LOCATION: (2740)..()
; NAME/KEY: misc_feature
; LOCATION: (800)..(808)
; OTHER INFORMATION: HAV tripeptide
; NAME/KEY: misc_feature
; LOCATION: (2225)..(2295)
; OTHER INFORMATION: transmembrane
; NAME/KEY: misc_structure
; LOCATION: (2296)..(2746)

```

; OTHER INFORMATION: cytoplasmic tail
; NAME/KEY: sig_peptide
; LOCATION: (109)..(558)
US-09-798-267-1

```

Alignment Scores:

Pred. No.:	4,298-232	Length:	2808
Score:	2349.50	Matches:	474
Percent Similarity:	67.24%	Conservative:	117
Best Local Similarity:	53.92%	Mismatches:	226
Query Match:	53.78%	Indels:	62
DB:	4	Gaps:	14

US-09-916-849A-1 (1-829) x US-09-798-267-1 (1-2808)

QY 4 ProArgGlyProLeuAlaSerLeuLeuLeuLeuValCysTrpLeuGlnCysAla-Al 23
Db 129 CTTCTCGGGCCCTGCCTGCTGCTAGGTCTCTCT-----TGCTTCTCAGGAGGGAG 182
QY 23 aserGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAlaGlyLysAl 43
Db 183 CCTCTCTCCTCT-----TTGACGCGAGAGCTACAGTTTCACGGTCCCCGGCGGCCACC 236
QY 43 agLuGlnGluProGlyGlnAlaLeuGlyLysVal---PheMetGlyCysProGlyGlnGl 62
Db 237 TGAGAAGAGCGCG-----CCTCTGGCGAGAGTGAATTTTCAAGATTGCACCGGTGCACA 290
QY 62 u----- 62
Db 291 AAGCACGACTATTTTCTTGACACCGATTCCGAAAGTGGGCACAGATGGTGTGATTACAGT 350
QY 63 -----ProAlaLeuPheSerThrAspAsnAspPheThrValIargAsnGlyLuth 80
Db 351 CAATAAGGCTCTACGGTTTTTAACCCACACAGATCCATTTCTTTGGTCTACGCTGGGACTT 410
QY 80 rValGlnGluArgArgSerLeuLys-----GluAArgAs 91
Db 411 CACCTACAGAAAAGTTTTTCCACCNAAGTCACGCTGAATACAGTGGGACACCCACCGCCC 470
QY 91 nPro-----LeuLysIllePhePro---Se 98
Db 471 CCCGCCCCCATCAGGCCTCCGTTTCTGGAATCCAAGCAGAATTCTCTCACATTTTCCCAACTC 530
QY 98 rlysArgileLeuArgArgHisLysArgAspTrpValValAlaproleIservValprogl 118
Db 531 CTCTCCTGGCCTCAGAAGACAGAGAGAGACTGGTATTCTCTCCATCAGCTGCCCAGA 590
QY 118 uAsnGlyLysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAsArgAs 138
Db 591 AAATGAAAAAGGCCCATTTCTTAAAAACCTGGTTTCAGATCAAAATCAAACAAAGACAAGA 650
QY 138 pThrLysIllePheTySerilleThrGlyProGlyValaaSPserProGluGlyValph 158
Db 651 AGGCAAGGTTTTCTACAGCATCACTGGCCAAGGAGCTGCACACACCCCTTGTGGTGTCTT 710
QY 158 eAlaValGluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluGluil 178
Db 711 TATTATTGAAAGAGAACAAGATGCGTGAAGGTGCAGAGCCCTCTCGATAGAGAAGCAT 770
QY 178 eAlalysTyrrGluLeuPheGlyHisAlaValSerGluAsnGlyValaserValGluasPr 198
Db 771 TGCCCATACACTCTCTCTCTCAGCTGTGTCTATCCACGGGAATGCAGTTTAGAGATCC 830
QY 198 oMetAsnIleSerillelvalThrAspGlnAsnAspHisLysProLysPheThrGlnAs 218
Db 831 AATGGAGATTTTGATCAGGTAACCGATCAGATGATGACAACAGCCCGAATTCAACCACAGA 890
QY 218 pThrPheArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValTh 238
Db 891 GGTCTTTTAAAGGGTCTGTCTATGGAAGGTGCTCTTCCAGGNACCTCTGTGATGAGAGTCA 950
QY 238 rAlaThrAspGluaspaspallalleTyThrTyAsnGlyValValAlaTySer-IleHi 258

Db	951	AGCCACACAGCGCGACGATGATGGAACACCTTAAATCGCCGCCATCGCTTACACCATCCCT	101
Qy	258	sSerGlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyTh	278
Db	1011	CAGCCAAAGATCTCTGAGCTCCCTGACAAAATATGTTTACCATTAAACAGGAACACAGAGT	1070
Qy	278	rIleSerValIleSerSerglyLeuAspArgGluLysValProGluTyrThrLeuThrI	298
Db	1071	CATCAGTGTGGTCACCACTGGGTGGACCGAGAGATTTCCTACGTATACCTCGTGGT	1130
Qy	298	eGlnAlaThrAspMetAspGlyAspGlySerThrThrAlaValAlaValGluI	318
Db	1131	TCAAGCTGCTGACCTTCAAGTGGGGTTAAGCACAAACAGCAACAGCTGTGTACAGT	1190
Qy	318	eLeuAspAlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValPr	338
Db	1191	CACTGACACCAACGATTAATCTCCGATCTTCAATCCCAACCCACGCTACAAGGTCAGGTGC	1250
Qy	338	oGluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAs	358
Db	1251	TGAGAACGAGGCTAAACGTCTGTATACCAACACTGAAAGTGACTGTGTGTATGCCCCAA	1310
Qy	358	nSerProAlaTirArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheTh	378
Db	1311	TACCCAGCGTGGAGGCTGTATACCACTATTG---AATGATGATGGTGGACAATTTGT	1367
Qy	378	rIleThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPh	398
Db	1368	CGTCACCAAAATCCAGTGAACACGATGCGATTTTGAACACAGCAAGGGCTTGATTT	1427
Qy	398	eGluAlaLysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValIle	418
Db	1428	TGAGGCCAAGCAGCAGTACATCTTACACGTAGCAGTGACGAATGTGGTACCTTTTGAGT	1487
Qy	418	uLysLeuProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaPr	438
Db	1488	CTCTCTACACACTCCACAGCCACCGTGCACCGTGGATGTCTGGATGTGAATGAAGGCC	1547
Qy	438	oValPheValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluPr	458
Db	1548	CATCTTTGTGCTCCTGAAAGAGAGTGGAGTGTCCGAGGACTTGGCGTGGCCAGGA	1607
Qy	458	oValCysValTyrThrAlaGluAspProAspLys---GluAsnGlnLysIleSerTyrAr	477
Db	1608	AATCATCTCTACACTGCCCGCAGGCGACACACATTTATGGAAACAGAAAATAACATATCG	1667
Qy	477	gIleLeuArgAspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGluValThrAl	497
Db	1668	GAITTTGGAGAGACACTCGCAACTGGCTTGAGATTAATCCGACACTGGTGCCATTTCAC	1727
Qy	497	aValGlyThrLeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMe	517
Db	1728	TCGGGCTGAGCTGGACAGGGAGATTTTGACGACGTGAAGAACAGCAGCTACACAGCCCT	1787
Qy	517	tValLeuAlaMetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuTh	537
Db	1788	AATCATAGCTACAGACAAATGGTGTCCAGTTGCTACTGGAACAGGACACTTCTGCTGAT	1847
Qy	537	rLeuIleAspValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAs	557
Db	1848	CCGTGCTGTGATGTGAATGACACGCCCCCATACCAGAACTCGAACTATATTTCTGTGCA	1907
Qy	557	nGlnSerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSe	577
Db	1908	GAGGAATCCAAAGCCTCAGGTCATAAAACATTATGATGACAGACCTTCTCCCAATACATC	1967
Qy	577	rProPheGlnAlaGlnLeuThrAspAspSerAspIleTyrTyrThrAlaGluValAsnGl	597
Db	1968	TCCCTTCACAGACGAACCTAACACACGGGGGAGTGGCCCACTGGACCATTCAGTACACGA	2027
Qy	597	u---GluGlyAspThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAs	616
Db	2028	CCCAACCCAAAGATCTATCATTTTCAAGCCAAAGATGGCTTTAGAGTGGGTGACTACAA	2087

REGISTRATION NUMBER: 36,637
REFERENCE/POCKET NUMBER: B0801/7023
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO.: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2808 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: liver
FEATURE:
NAME/KEY: CDS
LOCATION: 109..2745
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 109..558
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 559..2742
FEATURE:
NAME/KEY: cytoplasmic tail
LOCATION: 2296..2746
FEATURE:
NAME/KEY: HAV tripeptide
LOCATION: 800..808
FEATURE:
NAME/KEY: transmembrane
LOCATION: 2225..2295

PCT-US95-05518-1

Alignment Scores:

	Pred. No.:	Length:	Score:	Matches:
QY	4	4,296-232	2349.50	474
Db	129	Percent Similarity: 67.24%	Conservative: 117	
QY	23	Best Local Similarity: 53.92%	Mismatches: 226	
Db	183	Query Match: 53.78%	Indels: 52	
QY	43		Gaps: 14	
Db	237			
QY	62			
Db	291			
QY	63			
Db	351			
QY	80			
Db	411			
QY	91			

US-09-916-849A-1 (1-829) x PCT-US95-05518-1 (1-2808)

QY 4 ProArgGlyProLeuAlaSerLeuLeuValCysTrpLeuGlnCysAla-Al 23
Db 129 CCTCTCGGGCCCTGCTGCTGCTAGGTCTCTCT-----TGGCTCTCAGGACGGAG 182
QY 23 aserGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAlaGlyGlyAl 43
Db 183 CCCTCTCTCTCTGTT-----TTGACGCAGAGAGTCACGTTCCACGGTGCCCGGCCACC 236
QY 43 agluGlnGluProGlyGlnAlaLeuGlyLysVal---PheMetGlyCysProGlyGlnGl 62
Db 237 TGAGAAGAGAGCGCG-----CGTCTGGCAGAGTAGTAATTTTGAAGATTGCACGGTGCACA 290
QY 62 u----- 62
Db 291 AAGGACAGCTATTTTCTTGACACCGCATTCGAAAGTGGGCACAGATGGTGTATTACAGT 350
QY 63 -----ProAlaLeuPheSerThrAspAsnAspAspPheThrValArgAsnGlyGluTh 80
Db 351 CAAAAGGCCTCTACGGTTTTCATAACCAACACAGATCCAATTTCTTGCTCTACGCTGGGACTC 410
QY 80 rValGlnGluArgArgSerLeuLys-----GluArgAs 91
Db 411 CACCTACAGAAGATTTCACCAACAAAGTACAGCTCACGCTGAATACAGTGGGGACCACCGGCC 470
QY 91 nPro-----LeuLysIlePhePro---Se 98

616 pValHisSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaTh 636
2088 ATATCAATCTCAAGCTCATGGATACCAAGATAAGACCAAGTACCACTTAGAGGTGAG 2147
636 rValCysAspCysHisGlyHisValGluThrCys-----ProGlyProTrpLysGlyGI 654
2148 CGTGTGTGACTGTGAAGCGCGCGCGGCTCTGTAGGAAGGCACAGGCTGTGCAAGCAGG 2207
654 yPhe-----IleLeuProValLeuGlyAlaValLeuAlaLeuLeuPheLeuLe 670
2208 ATTGCAAAATCTCGCAATCTCGGGGATCTTGGAGGAATCTTCTCTTGTCTTATCTGAT 2267
670 uLeuValLeuLeuLeuValArgLysLysArgLysIleLysGluProLeuLeuLeuPr 690
2268 TCTGCTGCTCTGCTGTTTTCTTCGGAGGAGAGCGGTGTCAAAGAGCCCTTACTTGCCTCCC 2327
690 oGluAspAspThrArgAspAsnValPheTyrrGlyGluGluGlyGlyGlyGluGluAs 710
2328 AGAGGTATGACACCGCGGCAACGTTTATTACTATGATGAAGAAGAGCGCGGAGAGGA 2387
710 pGlnAspTyrrAspIleThrGlnLeuHisArgGlyLeuGluAlaArgProGluValValle 730
2388 CAGAGACTTGTACTTGAGCCAGCTGCACAGGGGCTCGAGCGCTGGCGTGAAGTGA-- 2445
730 uArgAsnAspValAlaProThrIleIleProThrProMetTyrrArgProArgProAlaAs 750
2446 -CGTAACGACGTTGCACCAACCTCATGAGTGTCCCCCGGTATCTTCCCGCCCTGCGCA 2504
750 nProAspGluIleGlyAsnPhelIlecluanLeuLysAlaAlaAsnThrAspProTh 770
2505 TCCCGATGAATATGAAAAATTTATTGATGAAAATCTGAAGCGCTGATCTGACCTGACCCAC 2564
770 rAlaProProTyrrAspThrLeuLeuValPheAspTyrrGluGlySerGlySerAspAlaAl 790
2565 AGCCCCGCTTATGATTCTCTGCTGCTGTTGACTATGATGAAGAAAGCGGTTCCGAAGCTGC 2624
790 aserLeuSerSerLeuThrSerSerAlaSerAspGlnAspGlnAspTyrrAspTyrrLeuAs 810
2625 TAGTCTGAGCTCCCTGAACTCTCAGAGTCAGAACAAAGACCGAGCTATGACTTACTTGAA 2684
810 nGluTPyrrGlySerArgPheLysLysLeuAlaAspMettTyrrGlyGlyGlyGluAsp 828
2685 CGAATGCGGCAATCGCTTCAGAAAGCTGCTGACATGTACGAGGCGCGGAGGAC 2739

RESULT 5
PCT-US95-05518-1
Sequence 1, Application PC/TUS9505518
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic
TITLE OF INVENTION: E-cadherin Interactions with T Lymphocytes
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05518
FILING DATE: herewith
PRIOR APPLICATION DATA:
FILING DATE: 08/23/919
FILING DATE: 3 May 1994
ATTORNEY/AGENT INFORMATION:
Name: Patrick J. Fitzmaurice
Firm: Fitzmaurice, Johnson, Finkelstein & Rothman LLP
Address: One Park Street, Suite 2000, New York, NY 10038
Phone: (212) 360-1000
Fax: (212) 360-1001
E-Mail: pfitz@fjfr.com

Db 471 CCGGCCCATCAGGCTCGTTTTCGGAATCCAAAGCAGAAATTCCTCACATTTTCCCAACTC 530
Qy 98 rlyArgilleLeuAArgHisIysArgAspTrpValValAlaProIleSerValProGl 118
Db 531 CTTCTCTGCTTCAGAGACAGAGAGAGACTGGGTATTTCTCCCATCAGCTGCCCA 590
Qy 118 uasnGlyLysGlyProPheProGlnArgLeuAenGlnLeuLysSerAsnLysAspArgAs 138
Db 591 AAATGAAAGGCCCATTTCTTAAACCTCGTTTCAGATCAATCCAAACAAGACAAGA 650
Qy 138 pThrIyIlePheThrSerIleThrGlyProGlyAlaAspSerProProGluGlyValph 158
Db 651 AGCGAGGTTTCTACAGCATCAGTGGCCAGAGGCTGACACCCCTGTGTGTCTT 710
Qy 158 eAlaValGluLysGluThrGlyTrpLeuLeuLysProLeuAspArgGluGluI 178
Db 711 TATTATTGAAGAGAAACAGGATGCTGAAGTGACAGAGCTCTGGATAGAACGCA 770
Qy 178 eAlaLysTrpGluLeuPheGlyHisAlaValSerGluAenGlyAlaSerValGluAspPr 198
Db 771 TGCCACATACACTCTCTCTCTCAGCTGTGTCTCAACCGGAATGCAATGAGGATCC 830
Qy 198 oMetAsnIleSerIleIleValThrAspGlnAsnAspHisIlyProLysPheThrGlnAs 218
Db 831 AATGGAGATTGATCACGGTAACCGATCAGATGACAGAACAGCCGGAATTCACCCAGGA 890
Qy 218 pThrPheArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValTh 238
Db 891 GGTCTTTAAGGGGTCTGTATGGAAGGTGCTCTTCCAGGAACCTCTGTGTGAGGTGAC 950
Qy 238 rAlaThrAspGluAspAlaIleThrThrTyAsnGlyValValAlaIleThrSerIleHi 258
Db 951 AGCCACAGACCGGACGATGATGTAACCTACAAATGCCGCCATCGCTTACACCATCT 1010
Qy 258 sSerGlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyTh 278
Db 1011 CAGCCAGATCTGAGTCCCTGTGACAAATAATGTTCCACATTAACAGGAACACAGAGT 1070
Qy 278 rIleSerValIleSerSerGlyLeuAspArgGluLysValProGluThrThrLeuThrI 298
Db 1071 CATCAGTGTGTACCATCGGCTGGACCGGAGAGATTTCCCTACGTATACCTCTGTGT 1130
Qy 298 eGlnAlaThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluI 318
Db 1131 TCAAGCTGCTGACCTTCAAGGTGAGGGTTAAGCACACACAGCAACACCTGTGTATCAGT 1190
Qy 318 eLeuAspAlaAsnAspAsnAlaProMetPheAspProGlnLysTrpGluAlaHisValPr 338
Db 1191 CACTGACACCAACAGATAATCTCCGATCTTCAATCCACACAGTACAAAGGTCAGGTGCC 1250
Qy 338 oGluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAs 358
Db 1251 TGAGAACGAGCTAACCTCGTAATCACCACTGAAAGTGACTGATGCTGATGCCCCCAA 1310
Qy 358 nSerProAlaTrpArgAlaThrTrpLeuIleMetGlyGlyAspAspGlyAspHisPheTh 378
Db 1311 TACCCACGCTGGAGGCTGTATACACATATTG---AATGATGATGGTGACAAATTGT 1367
Qy 378 rIleThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPh 398
Db 1368 CGTCACCAAAATCCATGTAACACAGATGCGCATTTTGAACACAGCAAGGCGCTTGATTT 1427
Qy 398 eGluAlaLysAsnGlnHisThrLeuTrpValGluValThrAsnGluAlaProPheValle 418
Db 1428 TGAGGCCAACAGCAGTACATTCTACAGTAGCAGTGACGAATGTGGTACCTTTTGGGT 1487
Qy 418 uLysLeuProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaPr 438
Db 1488 CTCTCTCACACCTCCACACCCACCGTACCGTGGATGTCTGGATGTGAATGAAGGCC 1547
Qy 438 oValPheValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluPr 458
Db 1548 CATCTTGTGCTCTTCTGAAAGAGAGTGGAAAGTGTCCGAGGACTTTGGCGCTGGGCCAGGA 1607

Qy 458 oValCysValTrpThrAlaGluAspProAspLys---GluAsnGlnLysIleSerTrpAr 477
Db 1608 AATCACATCTCTACTGCCAGGAGCCAGACACATTTATGAAACAGAAATAACATATCG 1667
Qy 477 gIleLeuArgAspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAl 497
Db 1668 GATTGGAGAGACACTCGCAACTGGCTGGAGATTAAATCCGAGCACTGGTCCCATTTCCAC 1727
Qy 497 aValGlyThrLeuAspArgGluAspGluGlnPheValArgAsnAsnIleTrpGluValMe 517
Db 1728 TCGGGCTGAGCTGCACAGGAGGATTTTGAGCAGCGTGAAGAACAGCAGCTACACAGCCCT 1787
Qy 517 tValLeuAlaMetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuTh 537
Db 1788 AATCATAGCTACAGACAATGGTTCTCCAGTTGCTACTGGAACAGGAGCACTTCTGTGTAT 1847
Qy 537 rLeuIleAspValAsnAspHisGlyProValProGluProArgGlnIleThrIleThrIleCysAs 557
Db 1848 CCTGTCTGATGTGAATGACACAGCCGCCCATACAGAACCTCGAACTATATTCTTCTGTGA 1907
Qy 557 nGlnSerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSe 577
Db 1908 GAGGAATCCAAAGCCCTCAGTCTATAAACATTCATGATGCAGACCTTCTCTCCAATACATC 1967
Qy 577 rProPheGlnAlaGlnLeuThrAspAspSerAspIleTrpThrAlaGluValAsnGl 597
Db 1968 TCCCTTCACAGCAGAACTAACACACGGCGAGTGCCTAGGACCTTTCAGTACACAGA 2027
Qy 597 u---GluGlyAspThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTrpAs 616
Db 2028 CCAACCCCAAGATCTATCATTTGAAGCCAAAGATGGCTTAGAGGTGGTGACTACAA 2087
Qy 616 pValHisLeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaTh 636
Db 2088 AATCAATCTCAAGCTCATGTAATAACAGAAATAAGACCAAGTACCACTTAGAGGTGAG 2147
Qy 636 rValCysAspCysHisGlyHisValGluThrCys-----ProGlyProTrpLysGlyGl 654
Db 2148 CGTGTGTGCTGTGAAGGGCGCCCGGCGTCTGTAGGAAGGCACAGCTGTGCAAGCAGG 2207
Qy 654 yPhe-----IleLeuProValLeuGlyAlaValLeuAlaLeuLeuPheLeu 670
Db 2208 ATTGCAAAATCTCGTCCATTTCTGGGATTTCTGGAGGAATTTCTTGTCTTAAATCTGAT 2267
Qy 670 uLeuValLeuLeuLeuValArgLysLysArgLysIleLysGluProLeuLeuLeuPr 690
Db 2268 TCTGTGCTCTTCTGTCTTCTTCGAGGAGAGCGGTGTCAAGAGAGCCCTTACTGCCCCC 2327
Qy 690 oGluAspAspThrArgAspAsnValPheTrpTrpGlyGluGlyGlyGlyGlyGluAs 710
Db 2328 AGAGGATGACACCCGGGACACGTTTATTATCTATGATCAAGAAGAGGAGCGGAGAGAGA 2387
Qy 710 pGlnAspTrpAspIleThrGlnLeuHisArgGlyLeuGluAlaArgProGluValValle 730
Db 2388 CCAGAGACTTTGACTTGAACCCAGCTGCACAGGCGCTGGACGCTCGGCTCGAAGTACT-- 2445
Qy 730 uArgAsnAspValAlaProThrIleProThrProMetTrpArgProArgProAlaAs 750
Db 2446 -CGTAACGAGCTGACCAACCCCTCATGAGTCTCCCCCGGTATCTTCCCCCGCTGCCAA 2504
Qy 750 nProAspGluIleGlyAsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAspProTh 770
Db 2505 TCCCGATGAAATGGAAATTTTATGTAATAATCTGAAGCGGCTGATCTGACCCCA 2564
Qy 770 rAlaProProTrpAspThrLeuLeuValPheAspTrpGluGlySerGlySerAspAlaAl 790
Db 2565 AGCCCCGCTTATGATCTCTGCTGCTTGTGACTATGAAGAAGCGGTTCGAAGCTGC 2624
Qy 790 aSerLeuSerSerLeuThrSerSerAlaSerAspGlnAspGlnAspTrpAspTrpLeuAs 810
Db 2625 TAGTCTGAGCTCCCTGAACTCTCAGAGTCAAGAAACAGCAGGACTATGACTACTTGA 2684

350 ThrValThrAspLeuAspAlaProAsnSerProAlaThrArgAlaThrThrLeuLeuMet 369
1254 ACGTGATGAGCCGAGATCAGCCCGCTCTCCAACTCGAATGCGTTTACCGCATCATC 1313
370 GlyGlyAspAspGlyAspHisPheThrThrHisProGluSerAsnGlnGlyLeu 389
1314 AGTGGGATCCATCCGGGCACTTCAGCTCGGCACAGACCCCGTAACCAACGAGGCGATG 1373
390 LeuThrThrArgLysGlyLeuAspPheGluAlaLysAsnGlnHisThrLeuThrValGlu 409
1374 GTCCCGTGGTGAAGCGATCGATACGAGCTCAACAGAGTTCATGCTCAGAGTATG 1433
410 ValThrAsnGluAlaProPheValLeuLysLeuPro-----ThrSerThrAlaThr 426
1434 GTGTCCAAACGAGGCGCCCTGGCCAGCGAATCCAGATGTCTTCCAGTCCACGCGAGGG 1493
427 IleValValHisValGluAspValAsnGluAlaProValPheValProProSerLysVal 446
1494 GTGACCATCTCCATCATGAGATCAACGAGGCTCCCTACTTCCCTCAAAACCAACGAGCTG 1553
447 ValGluValGlnGluGlyLeuProThrGlyGluProValCysValThrAlaGluAsp 466
1554 ATCCGCTGGAGGAGGGGTGCCCCCGGACCGCTGACCATCGTTTACGTGTGAGC 1613
467 ProAspLys-----GluAsnGlnLysLeuSerTyrArgIleLeuArgAspProAlaGlyTrp 485
1614 CCTACCGGTTTCATGACGAGCGGTGTGAGATCTCAAAAGCTGTCCAGACCCAGCGAGCTGG 1673
486 LeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThrLeuAspArgGluAsp 505
1674 CTGCACATCAATGCCAACCGGACGATCAACCGGTGGAGTGTGGACCGGAG--- 1730
506 GluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAlaMetAspAsnGlySer 525
1731 TCCCTCTACACCAAAACAACTGACGAGCCACCTTCTGGCGAGCTGACAAATGGGATA 1790
526 ProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeuAspValAsnAspHisGly 545
1791 CCCCCGCGCAGCGACCGGACCCCTCAGATCTATCTTCTTATGATCAACGACCAAGCC 1850
546 ProValProGluProArgGlnIleThrIleCysAsnGlnSerProValArgHisValLeu 565
1851 CCTGAGCTGTGCCCCAGGAGCGCAGATCTGC---GAGAGGCCCACTGACCGCATC 1907
566 AsnIleThr-----AspLysAspLeuSerProHisThrSerProPheGlnAlaGlnLeu 583
1908 AACATCAGCGCGCGCAGCTGACGTGACGCCCAACATCGGCCCTACGTCTTCGAGCTG 1967
584 ThrAspAspSerAspIleThrThrAlaGluValAsnGlu----- 597
1968 CCC-----TTGTCCCGCGCGCGTGGGAGAACTGGACCATCACCCGC 2012
598 ---GluGlyAspThrValValLeuSerLeuLys---LysPheLeuLysGlnAspThrTyr 615
2013 CTGAACGGTGACTATGCCCACTCAGCTTGGCATCTGTACCTGGAGCGCGGATGTAT 2072
616 AspValHisLeuSerLeuSerAspHisGlyAsnLys-----GluGlnLeuThrValIle 633
2073 GAGTCCCATCATCTGTCAGACTCTGGAACCCCTCCCTGTCACACAGCTCCATCATC 2132
634 ArgAlaThrValCysAspCys-----HisGlyHisValGluThrCysProGlyProTrp 651
2133 AAAGTCAAGGTGTGCCCATGTGATGACACGGGACTGACACCATTCATGGCGCAGTGGCA 2192
652 LysGlyGlyPheIleLeuProValLeuGlyAlaValLeuAlaLeuPheLeuLeuLeu 671
2193 GCGCTGGTCTG-----GGCAGCGGTGCCATCGTGGCCCATCTCATCTGCATCCTC 2243
672 ValLeuLeuLeuVal----- 681
2244 ATCTGCTGACCATGCTCTGCTGTTGTCATGTGGATGAAGCGGAGAGAGGAGCGC 2303
682 LysIleLysGluProLeuLeuLeuProGluAspThrArgAspAsnValPheTyrTyr 701

2304 CACAGAGAGCGTGTCTCATTTGACCCCGAGGACGAGTCCCGGAAAGATCTCTCAAGTAT 2363
702 GlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArgGly 721
2364 GACGAGGAGGCGGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2423
722 LeuGluAlaArgProGluVal-----ValLeuArgAsnAspValAla 735
2424 ---GAAGCCATGGGCGACGTCGCAAGAAAGCCCTCGGCGTGCCTGGATGAGCGG 2480
736 ProThrIleLeuProThrProMetTyr-----ArgProArgProAlaAsnProAspGlu 753
2481 CCG---GTGGCCCTGAGCCCGGATACCGATACCGATACCGATACCGATACCGATACCG 2537
754 IleGlyAsnPheIleLeuLeuLeuLeuLysAlaAlaAsnThrAspProThrAlaProPro 773
2538 ATCGGTGACTTCATCAATGAGGAGCTCCGCGTGTGTGACAAAGCAGCCCGGACCGCC 2597
774 TyrAspThrLeuLeuValPheAspTyrGlyGlySerGlySerAspAlaAlaSerLeuSer 793
2598 TATGACTCCCTGCTGTCTTTCGACTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2657
794 SerLeuThrSerSerAlaSerAspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGly 813
2658 TCCTCAACTCATCCAGTTCGCG---GACCAAGACTACGATTACCTCAACGACTGGGCG 2714
814 SerArgPheLysLysLeuAlaAspMetTyrGlyGlyGlyGluAspAsp 829
2715 CCAGATTCAAGAGCTGGCGACATGTATGAGGTGTGAAGAGAT 2762

RESULT 7

US-08-332-643-41
; Sequence 41, Application US/08332643
; Patent No. 5639634
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,643
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/872,643
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5639634and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3048 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

; TOPOLOGY: linear		
MOLECULE TYPE: CDNA		
US-08-332-643-41		
Alignment Scores:		
Pred. No.:	6,5e-157	3048
Score:	1621.50	371
Percent Similarity:	54.38%	Conservative: 138
Best Local Similarity:	39.64%	Mismatch: 296
Query Match:	37.11%	Indels: 132
DB:	1	Gaps: 22
US-09-916-849A-1 (1-829) x US-08-332-643-41 (1-3048)		
QY	4	ProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeu-----19
DB	19	CGCGGGCGCGCGGTGCTC-CTTCTGCTGCTCTCGCTCTCGCGCGCTCGCGGCCCAT 77
QY	20	-----GlnCysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThr 37
DB	78	AATGAGGATCTTACAACTAGAGACCTGCAAGCTGGTTCTCTGAAGATGATTACAG 137
QY	38	LeuGluAlaGlyGlyAlaGluGlnProGlyGlnAlaLeuGlyLysValPheMetGly 57
DB	138	GCATTATCTCCAAATATTCTAGAGGGGAAAGTACTTTCAGTCAAGTTTCAGCAGC 197
QY	58	CysProGlyGlnGluProAlaLeuPheSerThrAspAsnAspAspPheThrValArgAsn 77
DB	198	TGTGTGGGACCAAGGGGACAAATATGAGACCAACAGCATGGACTTCAAAGTTGGGGCA 257
QY	78	GlyGluThrValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePhePro 97
DB	258	GATGGGACAGTCTTCCGCCACCGGAGCTGCAG-----GTCCCC 296
QY	98	SerLysArgIle-----101
DB	297	TCCGAGAGGTGGCGTTTACCGTGTACTGATGGGACCCAGACAGCAGAGAGAAATGGAC 356
QY	101	-----101
DB	357	GCGTGTGTGCGGTGCTGTGTCGCGCCAGACCTCGTCCCGGCACTTGGACACAGCCGAG 416
QY	101	-----101
DB	417	AAAGAAAGAGGTGCTGTGCTGTGACCCCTCTCGCTCCGAGGACACCCCTGCTCGG 476
QY	102	-----LeuArgArgHisLysArgAspTrpValValAla 112
DB	477	TGGCCCCAGCAGCAGAACCGCAACCGGCTGAGGGCGGCAACCGGACTGGGTCTATCCCA 536
QY	113	ProfileSerValProGluAsnGlyLysGlyProPheProGlnArgLeuAsnGlnLeuLys 132
DB	537	CCCATCAACGTCCCGGAGAACTCGCGCGGCGCTTCCCGCAGCAGCTCGTGAAGATCCGG 596
QY	133	SerAsnLysAspArgAspThrLysIlePheTyrSerIleThrGlyProGlyAlaAspSer 152
DB	597	TCCGACAAAGACAATGACATCCCATCCGTCACGCTACAGCATCAGGGAGTGGTGGCGCAG 656
QY	153	ProProGluGlyValPheAlaValGluLysGluThrGlyTrpLeuLeuAsnLysPro 172
DB	657	CCCCCATGAGAGTCTTACAGATTAATCTCCATGTCGCGCGGAGTGTACTGCACAGGCC 716
QY	173	LeuAspArgGluGluIleAlaLysTyrGluLeuPheGlyHisAlaValSerGluAsnGly 192
DB	717	ATGGACCGGAGGAGCAGCCCTTTACCACTCCGAGCCACGCTGTGGATGAATGGC 776
QY	193	AlaSerValGluAspProMetAsnIleSerIleValThrAspGlnAsnAspHisLys 212
DB	777	AACAAGGTGGAGAACCCCATGCACTGTACATCTACGTCATGCAATGAATGACAAACAC 836
QY	213	ProLysPheThrGlnAspThrPheArgGlySerValLeuGluGlyValLeuProGlyThr 232
DB	837	CCTGAGTTTATCAACACAGGTCTTACAACTGCTCCGTGGAGGAGGGCTTCCAAAGCCAGGCACC 896
QY	233	SerValMetGlnValThrAlaThrAspGluAspAspAlaIleTyrThrTyrAsnGlyVal 252
DB	897	TACGTGATGACCATCACGCGCCACAGCTGCTGACGACGACAC--ACGGCCAAACGGGATG 953
QY	253	ValAlaTyrSerIleHisSerGlnGluProLysAspProHisAspLeuMetPheThrIle 272
DB	954	GTGCGGTACCGGATCGTGACCCAGACCCACAGAGCCGCTCCAGAAATATGTTCCACATC 1013
QY	273	HisArgSerThrGlyThrIleSerValIleSerSerGlyLeuAspArgGluLysValPro 292
DB	1014	AACAGCGAGACTGGAGATATCGTCAGTGGCGGTGCTGGGACCGGAGAGAAAGTTTCA 1073
QY	293	GluTyrThrLeuThrIleGlnAlaThrAspMetAspGlyAsp-----GlySerThr 309
DB	1074	CAGTACAGATCATCGTTTCAGGCCACAGATATGAGGAAATCTCAACTATGGCTCTCA 1133
QY	310	ThrThrAlaValAlaValIleLeuLeuAspAlaAsnAspAlaProMetPheAsp 329
DB	1134	AACACAGCCACAGCCATCATCAGGTGACAGATGTGAATGACAAACCCGTCAGAATTTACC 1193
QY	330	ProGlnLysTyrGluAlaHisValProGluAsnAlaValGlyHisGluValGlnArgLeu 349
DB	1194	GCCAGCACGTTTTCAGGGGAGGTCCCCGAAACACGCTGGAGACCGTGTGCGAAACCTC 1253
QY	350	ThrValThrAspLeuAspAlaProAsnSerProAlaTrpArgAlaThrTyrLeuIleMet 369
DB	1254	ACGGTGTGACCGAGATCAGCCCCACTCTCCAAATCGGAATGCGGTTTACCGCATCATC 1313
QY	370	GlyGlyAspAspGlyAspHisPheThrIleThrHisProGluSerAsnGlnGlyIle 389
DB	1314	AGTGGGATCCATCCGGGCACCTTCAGCTCCGACAGACCCCGTAACCAACAGAGGGCAT 1373
QY	390	LeuThrThrArgLysGlyLeuAspPheGluAlaLysAsnGlnHisThrLeuTyrValGlu 409
DB	1374	GTACCGTGTGAGGCGAGTGCATCAGCTCAACAGAGCTTTCATGCTGACAGTGTG 1433
QY	410	ValThrAsnGluAlaProPheValLeuLysLeuPro-----ThrSerThrAlaThr 426
DB	1434	GTGTCCACACGCGCCCTCGCCAGCGGAATCCAGATGTCTTCCAGTCCACGCGCAGG 1493
QY	427	IleValValHisValGluAspValAsnGluAlaProValPheValProSerLysVal 446
DB	1494	GTGACATCTCCATGATGACATCAACAGGCTCCCTACTTCCCTCAACACCAACAGCTG 1553
QY	447	ValGluValGlnGlyIleProThrGlyGluProValCysValTyrThrAlaGluAsp 466
DB	1554	ATCCGCTGGAGAGGGCGTCCCCCGGACCGGTGCTGACCCAGCTTTCAGCTGTGGAC 1613
QY	467	ProAspLys---GluAsnGlnLysIleSerTyrArgIleLeuArgAspProAlaGlyTrp 485
DB	1614	CCTGACCGGTTCATGACGAGCTGTGAGATACTCAAGCTGTGACACCCAGGAGCTGG 1673
QY	486	LeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThrLeuAspArgGluAsp 505
DB	1674	CTGCATCATCAATGCCACCAACGCGCAGATCACACCGTGGCAGTGTGGACCGTGA--- 1730
QY	506	GluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAlaMetAspAsnGlySer 525
DB	1731	TCCCTCTACACCAAAACACGCTACGAGGGCACCTTCTCGGACGCTGACATGGATA 1790
QY	526	ProProThrThrGlyThrGlyThrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 545
DB	1791	CCCCCGGCGAGCGGACCGGACCCCTCCAGATCTATCTTTCATTTGACATCAACCAACGCG 1850
QY	546	ProValProGluProArgGlnIleThrIleCysAsnGlnSerProValArgHisValLeu 565
DB	1851	CCTGAGCTGTGCGCCAAAGGAGGCGCAGATCTGC---GAGAGGGCCCAACCTGAAACGCAATC 1907
QY	566	AsnIleThr-----AspLysAspLeuSerProHisThrSerProPheGlnAlaGlnLeu 583
DB	1908	AACATCAGCGCGCGCGAGCTGACGTGACGCGCAGCCCAACATCGGCCCTTACCTCTCGAGCTG 1967

QY 584 ThrAspSerAspIleThrAlaGluValhAnGlu----- 597
Db 1968 CCC-----TTTGTCCGCGCGCGTGGGAAGACTGACCATCACCGC 2012
QY 598 ---GluGlyAspThrValValLeuSerLeuLys---LysPheLeuLysGlnAspThrTyr 615
Db 2013 CTGAACGGTGTACTATGCCCACTAGCTTGGCATCTGTACTTGGAGCGCGGATGAT 2072
QY 616 AspValHisLeuSerLeuSerAspHisGlyAsnLys-----GluGlnLeuThrValle 633
Db 2073 GAGTCCCATCATCTGTCAGAGACTCTGGAACCCCTCCCTGTCCACAGCTCCATCATC 2132
QY 634 ArgAlaThrValCysAspCys-----HisGlyHisValGluThrCysProGlyProTyr 651
Db 2133 AAAGTCAAGTGTGCCCATGTGATGACAAACGGGACTGCACCACTTGGCGAGTGGCA 2192
QY 652 LysGlyGlyPheIleLeuProValLeuGlyAlaValLeuAlaLeuPheLeuLeuLeu 671
Db 2193 CGCGCTGTCTG-----GGCACCGGTGCCATCTGTGGCCATCTCATCTGCTCCTC 2243
QY 672 ValLeuLeuLeuVal-----ArgLysLysArg 681
Db 2244 ATCTGTGTGACCATGCTCTGCTGTTGTCATGTGGATGAAGCGCGGAGAGAGGAGCGC 2303
QY 682 LysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyrTyr 701
Db 2304 CACAGGAGCAGCTGCTCATTCACCGGAGGAGCAGCTCCGGAAGAGATCCTCAAGTAT 2363
QY 702 GlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArgGly 721
Db 2364 GACAGGAGCGGTGGGAGGAGGAGCAGGACTACGACTACGACTCAGCCAGCTGCAGAGCGC 2423
QY 722 LeuGluAlaArgProGluVal-----ValLeuArgAspAspValAla 735
Db 2424 ---GAAGCATGGGCGACGTGCCAAGAAAGCCCTCGCGTGGTGGTGGTGGTGGTGG 2480
QY 736 ProThrIleLeuProThrProMetTyr-----ArgProArgProAlaAsnProAspGlu 753
Db 2481 CCG---GTGGGCCCTGAGCCCCAGTACCGATAGGCCCATGTGGCGCCAGCCAGCGAC 2537
QY 754 IleGlyAsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAspProThrAlaProPro 773
Db 2538 ATCGTGCATTCATCAATAGGAGTCTCGCGCTGCTGACCAACGCCACCGCACCCCCC 2597
QY 774 TyrAspThrLeuLeuValPheAspTyrGluLysSerGlySerAspAlaAlaSerLeuSer 793
Db 2598 TATGACTCCCTGTGTGTCCTTCGACTACGAGGGAGCGGCTCCACCGCAGGCTCCGTGAGC 2657
QY 794 SerLeuThrSerSerAlaSerAspGlnAspTyrAspTyrAspTyrLeuAsnGluTyrGly 813
Db 2658 TCCTGTGACTATCCAGTTCGGG---GACCAAGACTACGATTACCTCAAGCTGGGGC 2714
QY 814 SerArgPheLysLysLeuAlaAspMetTyrGlyGlyGlyGluAspAsp 829
Db 2715 CCCAGATTCAAGAGCTGGCGGACATGTATGGAGGTGGTGAAGAGGAT 2762

RESULT 8

US-08-332-638-47

; Sequence 47, Application US/08332638

; Patent No. 5646250

; GENERAL INFORMATION:

; APPLICANT: Suzuki, Shintaro

; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS

; NUMBER OF SEQUENCES: 62

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

; ADDRESSEE: Borun

; STREET: 6300 Sears Tower, 233 S. Wacker Drive

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,638
; FILING DATE: 01-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,643
; FILING DATE: 17 APR 1992
; APPLICATION NUMBER: US/08/049,460
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5646250 and, Greta B.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31340
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3048 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-332-638-47

Alignment Scores:
Pred. No.: 6,5e-157 Length: 3048
Score: 1621.50 Matches: 371
Percent Similarity: 54.38% Conservative: 138
Best Local Similarity: 39.64% Mismatches: 296
Query Match: 37.11% Indels: 132
DB: 1 Gaps: 22

US-09-916-849A-1 (1-825) x US-08-332-638-47 (1-3048)

QY 4 ProArgGlyProLeuAlaSerLeuLeuGlnValCysTyrLeu----- 19
Db 19 CGCGGGCGCGCGGTGCTC-CTTCTGTGCTCTCGCTCTCGGGCGCGTCCGGGCCCAT 77
QY 20 -----GlnCysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThr 37
Db 78 AATGAGGATCTTACAACTAGAGAGACCTCAAGGCTGGTCTCTGAAGATGATTACAGC 137
QY 38 LeuGluAlaGlyGlyValaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGly 57
Db 138 GCATTAACTCCCAAAATATTCTAGAGGGGAAAAGCTACTTCAAGTCAAGTTCAGCAGC 197
QY 58 CysProGlyGlnGluProAlaLeuPheSerThrAspAsnAspAspPheThrValArgAsn 77
Db 198 TGTGTGGGACCAAGGGGACACAATATGAGACCAACAGCATGAGCTTCAAGTTGGGCA 257
QY 78 GlyGluThrValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePhePro 97
Db 258 GATGGGACAGTCTTCGCCACCGGAGCTGCAG-----GTCCCC 296
QY 98 SerLysArgIle----- 101
Db 297 TCCGAGCAGGTGGCTTCCCGTGTGACTGCTGAGGACCCAGACAGAGAAATGGAC 356
QY 101 ----- 101
Db 357 GCGGTGTGCGGTGCTGTGGTGGCCAGACCTCGTCCCGCAGCTCTGGACACAGCGCGAG 416
QY 101 ----- 101
Db 417 AAAGGAAAGAGGTGCTGGTCTGGACCCCTCTCCGCTCCGAGGACACACCCTGTCGCG 476
QY 102 -----LeuArgArgHisLysArgAspTyrValValAla 112

```

Db      477 TGGCCCGACGACAGACCGCCACCGGCTGAGCGGGCGCAACCGGACTGGGTCAATCCCA 536
Qy      113 ProIleSerValProGluAenGlyLysGlyProPheProGlnArgIleuAenGlnLeuLys 132
Db      537 CCCATCAACGTGCCGAGAACTCGCGGGCGCTTCCCGCAGCAGCTCGTGAAGATCCGG 596
Qy      133 SerAsnLysAspArgAspThrLysIlePhePheTyrSerIleThrGlyProGlyAlaAspSer 152
Db      597 TCCGACAAAGACAAATGACATCCCATCCGTCACGATACCGGAGTGGGTGCGACCCAG 656
Qy      153 ProProGluGlyValPheAlaValGluLysGluThrGlyTrpLeuLeuLeuAenLysPro 172
Db      657 CCCCCCATGGAGGTCTTCAGCATTAATCTCCATGTCGCGCGCGATGTACGTCAACAGGCC 716
Qy      173 LeuAspArgGluGluIleAlaLysTyrGluLeuPheGlyHisAlaValSerGluAenGly 192
Db      717 ATGACCGGGAGGACGCGCTCTTACACCTCCGAGCCCGCATGTCGTCGACATGAATGGC 776
Qy      193 AlaSerValGluAspProMetAsnIleSerIleValThrAspGlnAsnAspHisLys 212
Db      777 AACAGGTGGAGAACCCCATCGACTGTACATCTACGTATCGACATGAATGACAAACAC 836
Qy      213 ProLysPheThrGlnAspThrPheArgGlySerValLeuGluGlyValLeuProGlyThr 232
Db      837 CCTGAGTTTCATCAACACCGGTCTACAACCTCGCTGGAGCGGCTCCAAAGCCAGGCACC 896
Qy      233 SerValMetGlnValThrAlaThrAspGluAspAlaIleTyrThrTyrAsnGlyVal 252
Db      897 TAGGTGATGACATCACCGCCACAGATGCTGACGACAGACACC---ACGGCCACCGGATG 953
Qy      253 ValAlaTyrSerIleHisSerGlnGluProLysAspProHisAspLeuMetPheThrIle 272
Db      954 GTGGGTACCGGATCGTGACAGCCAGACCCACAGAGCCCGTCCAGAGATATGTTCAACCATC 1013
Qy      273 HisArgSerThrGlyThrIleSerValIleSerValIleSerGlyLeuAspArgGluLysValPro 292
Db      1014 AACAGCGAGATCGAGATATCGTCACAGTGGCGGCTGGTGGGACCGAGAGAAAGTTTCAG 1073
Qy      293 GluTyrThrLeuThrIleGlnAlaThrAspMetAspGlyAsp-----GlySerThr 309
Db      1074 CAGTACACAGTCATCGTTCCGCGCACAGATATGAGAGAAATCTCAACTATGCGCTCTCA 1133
Qy      310 ThrThrAlaValAlaValGluIleLeuAspAlaAsnAspAlaProMetPheAsp 329
Db      1134 AACACAGCCACGACCATCACCGTGACAGATGTGAATGACACCCCGTCAGAAATTAC 1193
Qy      330 ProGlnLysTyrGluAlaHisValProGluAsnAlaValGlyHisGluValGlnArgLeu 349
Db      1194 GCCAGCAGTTTGCAGGGAGGTCCCGAAACAGCGTGGAGACCGTGGTCCGCAACCTC 1253
Qy      350 ThrValThrAspLeuAspAlaProAsnSerProAlaTrpArgAlaThrTyrLeuIleMet 369
Db      1254 ACGGTGATGGACCGAGATCAGCCCACTCTCCAAACTGGAATGCCGTTCACCGCATCATC 1313
Qy      370 GlyGlyAspAspGlyAspHisPheThrIleThrHisProGluSerAsnGlnGlyIle 389
Db      1314 AGTGGGATCCATCCGGGCATTTAGCGTCGCGACAGCCCGTAAACACAGAGGGCATG 1373
Qy      390 LeuThrThrArgLysGlyLeuAspPheGluAlaLysAsnGlnHisThrLeuTyrValGlu 409
Db      1374 GTCACCGTGGTGAAGGCAGTCGACTACGAGCTCAACAGAGCTTTCATGCTGACAGTGTATG 1433
Qy      410 ValThrAsnGluAlaProPheValLeuLysLeuPro-----ThrSerThrAlaThr 426
Db      1434 GTGTCCAAACAGGCGCCCTCGCCACGGAATCCAGATGTCCTTCCAGTCCACCGGCGGG 1493
Qy      427 IleValValHisValGluAspValAsnGluAlaProValPheValProProSerLysVal 446
Db      1494 GTGACCATCTCCATCATGACATCAACGAGGCTCCCTACTTCCCTCCAAACCAACAGCTG 1553
Qy      447 ValGluValGlnGluGlyIleProThrGlyGluProValCysValTyrThrAlaGluAsp 466

```

```

1554 ATCCGCTCGAGGAGGCGTGTGCTCCCGCGCACCGTGTGACCAAGTCTTACGTGTGGAC 1613
Qy      467 ProAspLys---GluAsnGlnLysIleSerTyrArgIleLeuArgAspProAlaGlyTrp 485
Db      1614 CTTGACCGGTTCATGTCACGAGGTGTGAGATACTCAAGCTGTGACAGCCAGCGAGTGG 1673
Qy      486 LeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThrLeuAspArgGluAsp 505
Db      1674 CTGCACATCAATGCCCAACGGCCAGATCACACGGGTGGCAGTGTGGACCGTGAG--- 1730
Qy      506 GluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAlaMetAspAsnGlySer 525
Db      1731 TCCCTCTTACCAACAAACAGCTACAGGCCACCTCCAGATCTATCTATTGACATCAACGACCA 1790
Qy      526 ProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAspValAsnAspHisGly 545
Db      1791 CCCCCGCGCAGCGGCACCGGACCCCTCCAGATCTATCTATTGACATCAACGACCAACGCC 1850
Qy      546 ProValProGluProArgGlnIleThrIleCysAsnGlnSerProValArgHisValLeu 565
Db      1851 COTGAGCTGCTCCCGACGAGGCGGACATCTGC---GAGAGGCCCAACCTGAGACGCCATC 1907
Qy      566 AsnIleThr-----AspLysAspLeuSerProHisThrSerProPheGlnAlaGlnLeu 583
Db      1908 AACATCAGCGCGCGCGACGCTGACGTGCGACCCCAACATCGGCCCTACGTCTTCGAGCTG 1967
Qy      584 ThrAspAspSerAspIleTyrTrpThrAlaGluValAenGlu----- 597
Db      1968 CCC-----TTTGTCCCGCGCGCGCTGCGGAAGAACTGGACCATCACCGCG 2012
Qy      598 ---GluGlyAspThrValValLeuSerLeuLys---LysPheLeuLysGlnAspThrTyr 615
Db      2013 CTGAACGTGATGATGCCCACTAGCTTGGCATCTGTACTTGGAGCGCGGATGAT 2072
Qy      616 AspValHisLeuSerLeuSerAspHisGlyAsnLys-----GluGlnLeuThrValIle 633
Db      2073 GACGTCCTCCATCATCGTCACAGACTCTCGAAACCTCCCTCTGCCAACACACGTCATCATC 2132
Qy      634 ArgAlaThrValCysAspCys-----HisGlyHisValGluThrCysProGlyProTrp 651
Db      2133 AAAGTCAGGTGTGCCCATGTGATGACAAACGGGAGTCAACACCATTTGGCGCAGTGCA 2192
Qy      652 LysGlyGlyPheIleLeuProValLeuGlyAlaValLeuAlaLeuLeuPheLeuLeu 671
Db      2193 GCGGCTGGTCTG-----GGCACCGTGCATCGTGGCCATCTCATCTGCATCTC 2243
Qy      672 ValLeuLeuLeuVal-----ArgLysLysArg 681
Db      2244 ATCTCTGTGACCATGTCCTGTGTTGTGATGATGAAGCGGCGAGAGAGAGCGC 2303
Qy      682 LysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyrTyr 701
Db      2304 CACACGAGCAGCTGCTCATTCACCCCGAGCAGCGTCCGCGAAAGATCTCTCAAGTAT 2363
Qy      702 GlyGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnIleHisArgGly 721
Db      2364 GACGAGGAGGCGGTGGCGAGGAGGACGAGGACTACGACCTCAGCCAGCTGCGAGCGCG 2423
Qy      722 LeuGluAlaArgProGluVal-----ValLeuArgAsnAspValAla 735
Db      2424 ---GAAGCCATGGGCACGTGCCAAGCAAGCCCTTGGCGTGGTGGTGGATGAGCGG 2480
Qy      736 ProThrIleProThrProMetTyr-----ArgProArgProAlaAsnProAspGlu 753
Db      2481 CCG---GTGGGCGCTGAGCCCGCAGTACCGCATCGGCGCATGGTCCCGCAGCGCGAC 2537
Qy      754 IleGlyAsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAspProThrAlaProPro 773
Db      2538 ATCGGTGATCTCATCAATGAGGACTCCGCGCTGTGTACACAGCCACCCCGCAGCCGCC 2597
Qy      774 TyrAspThrLeuLeuValPheAspTyrGluGlySerGlySerAspAlaAlaSerLeuSer 793
Db      2598 TATGACTCCCTGCTGTCTTCTTCTGACTACGAGGGGAGCGGCTCCACCGCAGGCTCCGTCAGC 2657

```

QY 794 SerLeuThrSerSerAlaSerAspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGly 813
|||||
Db 2658 TCCTGAACTCATCTCCGG--GACCAAGACTACGTTACCTCAAGACTGGGC 2714
QY 814 SerArgPheLysLeuAlaAspMetTyrGlyGlyGluAsp 829
|||||
Db 2715 CCCAGATTCAAGAGCTGGCGGACATGATGGAGGTGGTGAAGAGAT 2762

RESULT 9

US-08-188-228-61
; Sequence 61 Application US/08188228
; Patent No. 5597725
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,228
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/049,460
; FILING DATE: 19 APR 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,643
; FILING DATE: 17 APR 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5597725and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31340
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2690 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-188-228-61

Alignment Scores:
Pred. No.: 1 69e-95 Length: 2690
Score: 1025.50 Matches: 266
Percent Similarity: 51.89% Conservative: 119
Best Local Similarity: 35.85% Mismatches: 284
Query Match: 23.47% Indels: 73
DB: 1 Gaps: 22

US-09-916-849A-1 (1-829) x US-08-188-228-61 (1-2690)

QY 4 ProArgGlyProLeuAla----SerLeuLeuLeuGlnValCysTrpLeuGlnCysAla 22
|||||
Db 451 CCGAGAACTCGCTCGTTCTGTGGTTCTCTCCAGGTGCTGCTGAACATCTGCA 510
|||||
QY 23 AlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAlaGlyGly 42
|||||
|||||

Db 511 GAAGATTTGGACTGCCTCTGGATTTTCAGCAGAAAGTGTTCATATCAATCAGCCAGCT 570
QY 43 AlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGlyGlnGlu 62
|||||
Db 571 GAATTCATTTGAGCAGCAGTCAATTCCTAACTTGACCTTCAGTACTGTAAAGGAACGAC 630
QY 63 ProAlaLeuPheSerThrAspAsnAspPheThrVal----- 75
|||||
Db 631 AAGCTACGCTATGAGGTCTCGAGCCCATACTTCAAGGTGAACAGCGATGGGGCTTAGTT 690
QY 76 -----ArgAsnGlyGluThrValGlnGluArgArgSerLeuLysGluArgAsnPro--- 92
|||||
Db 691 GCTCTGAGAAACATAAATCTGAGTGGGCAAACTGTGTTCGTCCTCATGACAGGCCCAT 750
QY 93 -----LeuLysIlePheProSerLysArgIle----- 101
|||||
Db 751 GCGGAAGATATGCGCAGAACTCTGTATTCGGGGGAAAGACATCCAGGGTCTCTTCGAG 810
QY 102 -----LeuArgArgHisLysArgHisLysArgAspTrpValVal 111
|||||
Db 811 GATATATTTAAATTTGCAAGAACTTCTCTGTGCCAAGACAAAGAGGTCCATTTGGTGA 870
QY 112 AlaProIleSerValProGluAsnGlyLysGlyProPheProGlnArgLeuAsnGlnLeu 131
|||||
Db 871 TCTCCCATTTTAATTCAGAGATCAGACAGACGCTTTCCCAAGAGATGTTGGCAAGTA 930
QY 132 LysSerAsnLysAspArgAspThrLysIlePheTyrSerIleThrGlyProGlyAlaAsp 151
|||||
Db 931 ---GTCGATAGTACAGCGCCAGAAAGTCCAGTTCGGGCTCACTGGAAGAGTGGAT 987
QY 152 SerProGluGlyValPheAlaValGluLysGluThrGlyTrpLeuLeuAsnLys 171
|||||
Db 988 CAAGAGCCTTAAGGAATTTTCAATCAATGAGACACAGGAGGCGTCTCCGTGACAGG 1047
QY 172 ProLeuAspArgGluGluIleAlaLysTyrGluLeuPheGlyHisAlaValSerGluAsn 191
|||||
Db 1048 ACCTTGGACAGAGAAGTAATCGCTCTTTATCACTATTGTTGGAGACCACCTGATCTCA 1107
QY 192 GlyAlaSerValGluAspProMetAsnIleSerIleValThrAspGlnAsnAspHis 211
|||||
Db 1108 GCGAAACTCTCGAGGGCGCGTCTCTGGAAGTCATTGTGATTGATGATGATGATGATGAT 1167
QY 212 LysProLysPheThrGlnAspThrPheArgGlySerValLeuGluGlyValLeuProGly 231
|||||
Db 1168 CGACCGATCTTCGGGAAGGCCCTTACATCGGCCACGTCATGGAAGGGTCAACCCACAGG 1227
QY 232 ThrSerValMetGlnValThrAlaThrAspGluAspAlaIleTyrThrTyrAsnGly 251
|||||
Db 1228 ACCACAGTATGCGGATGACAGCCTTTGATGATGATGATGATGATGATGATGATGATG 1284
QY 252 ValValAlaTyrSerIleHisSerGlnGluProLysAspProHisAspLeuMetPheThr 271
|||||
Db 1285 CTCCTGGCGTATATATTCGTCACAGAGCGCTGCAAGCCATCTCCCAACATGTTCTAC 1344
QY 272 IleHisArgSerThrGlyThrIle---SerValIleSerSerGly---LeuAspArgGlu 289
|||||
Db 1345 ATCGATCTCGAAGAGGAGACATTGCTACTGTGTGTCTCACCTCGCTGCGTGGACGAGAG 1404
QY 290 LysVal-----ProGluTyrThrLeuThrIleGlnAlaThrAspMetAspGly----- 305
|||||
Db 1405 ACTCTGGAAATCCCAAGTATGAACTGATCATCGAGGCTCAAGATATGCTGGATGGAT 1464
QY 306 AspGlySerThrThrAlaValAlaValGluIleLeuAspAlaAsnAspAsnAla 325
|||||
Db 1465 GTTGGATTAAACAGGCACGCCACAGCCAGCATGATCGATGACAAAAATGATCACTCA 1524
QY 326 ProMetPheAspProGlnLysTyrGluAlaHisValProGluAsnAlaValGlyHisGlu 345
|||||
Db 1525 CCAAAATTCACCAAGAAAGAGTTTCAAGCCACAGTCGAGGAGGAGGCTGTGGGA---GTT 1581
QY 346 ValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerProAlaTrpArgAlaThr 365
|||||
Db 1582 ATTGTCAATTTGACAGTTGAAGATGAAGTATGACCCACACAGGTGATGAGGGGCTGCC 1641

366 TyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThrHisProGluSer 385
1642 TACACCATCATCAACGGAACCCCGGCGAGAGCTTTGAATCCACACCAACCTCAAC 1701
386 AsnGlnGlyIleLeuThrArgLysGlyLeuAspPheGluAlaLysAsnGlnHisThr 405
1702 AACGAGGAGATGCTTCTGTGTCACCAACCATGACTATGAAATTTCTGCTTCCACACC 1761
406 LeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu-----ProThr 422
1762 CTGCTGATCAAGTGGAAATGAAGACCCACTCGTACCCGACGCTCTCTACGCGCCGACG 1821
423 SerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPheValPro 442
1822 TCCACAGCCACCGTCCACATCATCTGCTGATGTCACAGGCGCCGAGTCTTCTACCCA 1881
443 ProSerLysValValGluValGlnGlyIleProThrGlyGluProValCysValTyr 462
1882 GACCCCATGATGGTACCGACGAGGAGCTCTCTGTGGCAGCGTGTCTGACAGTG 1941
463 ThrAlaGluAspProAspLys---GluAsnGlnLysIleSerTyrArgIleLeuArgAsp 481
1942 AATGCCACGACCCCGACTCCCTGCACATCAACCATCAGGTATCTGTTTCAAGGAC 2001
482 ProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThrLeu 501
2002 CCACAGGTGGCTGAATATTAAATCCATCAATGGGACTTTCACACCCAGCTGTGCTG 2061
502 AspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAlaMet 521
2062 GACCGTGAG---TCCCATTTGTGCACACAGCGGTGTACATGCTCTCTCTCTGCAATT 2118
522 AspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuLeuLeuLeuLeu 541
2119 GACAGTGGCAACCTCCCGTACGGGCACTGGGACTTGTCTGATACCCCTGGAGGACGTG 2178
542 AsnAspHisGlyProValProGluProArgGlnIleThrIleCys-----Asn 557
2179 AATGACAATGCCCGTTCATTTACCCACACAGTAGCTGAAGTCTGTGATGATGCAAAAAC 2238
558 GlnSerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSer 577
2239 CTCAGTGTAGTC-----ATTTCGGAGCATCAGATAAGATCTTCACCCGATACAGAT 2292
578 ProPheGlnAlaGlnLeuThrAspAspSer-----AspIleTyrTrpThrAlaGluVal 595
2293 CTTTCAATTTGAATCCACAAACAGCTTCTCTGATAAAGTCTGG-----AAGATC 2346
596 AsnGluGluGlyAspThr---ValValLeuSerLeuLysLysPheLeuLysGlnAspThr 614
2347 TCCAGATCAACAATACACACGCGCTCGTAAAGCTTCTTCAAAATCTGAAACAAAGCAAC 2406
615 TyrAspValHisLeuSerLeuSerAspHisGlyAsn-----LysGluGlnLeuThrVal 632
2407 TACACCTGCCCATCTGCTGACAGATTCAGGNAACCCCATGACGATATACAGAT 2466
633 IleArgAlaThrValCysAspCysHisGlyHisValGluThrCysProGlyProTrpLys 652
2467 CTCAGGGTACAAAGTGTGCTCTCGAGGAATTCAAAGTGGACTGCAACGCGGGGGGCC 2526
653 GlyGlyPheIleLeuProValLeuGlyAlaValLeuAlaLeuPheLeuLeuVal 672
2527 CTGGCTTCACCGTGC---TCAGTCTGCTCTCTACGCTCTCTACGCTTAGCTTGT 2580
673 LeuLeuLeuLeuVal-----ArgLysLys 680
2581 CTGTGAGAACTCTCTGACGCTGTGAAGCTTGACTCCCAAGTTTCCATGCAACAGGAAAAA 2640
681 ArgLys 682
2641 AAAAAA 2646

RESULT 10
US-08-332-643-55
Sequence 55, Application US/08332643
Patent No. 5639634
GENERAL INFORMATION:
APPLICANT: Suzuki, Shintaro
TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSER: Bicknell
STREET: Two First National Plaza, 20 South Clark
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/332,643
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/872,643
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: No. 5639634and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/30795
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 2690 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-332-643-55
Alignment Scores:
Pred. No.: 1,698-95 Length: 2690
Score: 1025.50 Matches: 266
Percent Similarity: 51.89% Conservative: 119
Best Local Similarity: 35.85% Mismatches: 284
Query Match: 23.47% Indels: 73
DB: 1 Gaps: 22
US-09-916-849A-1 (1-829) x US-08-332-643-55 (1-2690)
QY 4 ProArgGlyProLeuAla---SerLeuLeuLeuLeuGlnValCysTrpLeuGlnCysAla 22
Db 451 CCGAAGTAATCCGCTCGTCTGTGGCTTCTCTGTCGCCAGGTCGCTCAACATCTGCA 510
QY 23 AlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAlaGlyGly 42
Db 511 GAAGATTGGACTGCACCTCGGATTTTCAGCAGAAAGTGTCCATATCAATCAGCCAGCT 570
QY 43 AlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGlyGlnGlu 62
Db 571 GAATTCATGTAGGACGACCATCAATTCATAACTTGACCTTCAGTGTGTAAGGAAACGAC 630
QY 63 ProAlaLeuPheSerThrAspAsnAspPheThrVal----- 75
Db 631 AAGCTACGCTATGAGGTCTCGAGGCCATACCTTCAAGGTGAACGATGGCGGCTTAGTT 690
QY 76 -----ArgAsnGlyGluThrValGlnGluArgSerLeuLysGluArgAsnPro---- 92

Db	691	GCTCTGAGAACATAACTGCAGTGGGCAAAACTGTGTCTGCATGCACGGACCCCCCAAT	750
Qy	93	-----LeuLysIlePheProSerLysArgIle-----	101
Db	751	CGCGAAGATATGCGCAGAACTCGTCATGTTCGGGGGNAAGACATCCAGGCTCCTTGCAG	810
Qy	102	-----LeuArgArgHisLysArgAspTrpValVal	111
Db	811	GATATATTTAAATTTGCAAGAACTCTCTCTGTCGCCAGACAAAAGAGGTCCATGTGGTGA	870
Qy	112	AlaProIleSerValProGluAsnGlyLysGlyProPheProGlnArgIleAsnGlnLeu	131
Db	871	TCCTCCCATTTAATTCAGAGAAATCAGACAGACGCCITTCACAGAGATGTTCGCAAGTTA	930
Qy	132	LysSerAsnLysAspArgAspThrLysIlePheTySerIleThrGlyProGlyAlaAsp	151
Db	931	---GTCGATAGTCACAGCGCCAGAAAGGTCGAAGTTCGGGCTCACTGCGAAAGGAGTGGAT	987
Qy	152	SerProProGluGlyValPheAlaValGluLysGluThrGlyTrpLeuLeuLeuAsnLys	171
Db	988	CAAGAGCCTTAAGGAATTTTCAGAAATCAATGAGACACAGGAGCGTCTCCGTGACACGG	1047
Qy	172	ProLeuAspArgGluGluIleAlaLysTyTrpGluLeuPheGlyHisAlaValSerGluAsn	191
Db	1048	ACCTTGACAGAGAAAGTAATCGCTGTTTATCAACTATTCTGGAGACCACTGATGTCAAT	1107
Qy	192	GlyAlaSerValGluAspProMetAsnIleSerIleValThrAspGlnAsnAspHis	211
Db	1108	GGCAAAACTCTCAGAGGGCGCGTCTCGAAGTCATTGTGATGATCAGAAATGCAAC	1167
Qy	212	LysProLysPheThrGlnAspThrPheArgGlySerValLeuGluGlyValLeuProGly	231
Db	1168	CGACCGATCTTCGGGAAGGCCCTACATCGGCCAGCTCATGGAAGGGTCACCCACAGGC	1227
Qy	232	ThrSerValMetGlnValThrAlaThrAspGluAspAlaIleTyThrTyAsnGly	251
Db	1228	ACCACAGTGTGGGATGAGACGCTTTGATGTCAGATGACCCAGCC---ACCGATAATGCC	1284
Qy	252	ValValAlaTySerIleHisSerGlnGluProLysAspProHisAspLeuMetPheThr	271
Db	1285	CTCTCGGGTATAATATCCGTCAACAGACGCGCTGACAGGCATCTCCACATGTTCTAC	1344
Qy	272	IleHisArgSerThrGlyThrIle---SerValIleSerSerGly---LeuAspArgGlu	289
Db	1345	ATCGATCCTCAGAAAGAGACATGTCACTGTTGTGTCACTCGGTGCTGGACCGAGAG	1404
Qy	290	LysVal-----ProGluTyThrLeuThrIleGlnAlaThrAspMetAspGly-----	305
Db	1405	ACTCTGGAATAATCCCAAGATGAACATGATCATCGAGGCTCAAGATATGGCTGACTGGAT	1464
Qy	306	AspGlySerThrThrAlaValAlaValGluIleLeuAspAlaAsnAspAsnAla	325
Db	1465	GTTGGATTAACAGGCACGGCCACAGCCACGATCATGATCGATGACAAAAATGATCACTCA	1524
Qy	326	ProMetPheAspProGlnLysTyTrpGluAlaHisValProGluAsnAlaValGlyHisGlu	345
Db	1525	CCAAAAATCCCAAGAAAGAGTTTCAAGCCACAGTCGAGGAAGAGCTGGGGA---GTT	1581
Qy	346	ValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerProAlaTrpArgAlaThr	365
Db	1582	ATTGTCATTTGACAGTTGAAGATAAGATGACCCCAACAGCGTGCATGAGGGCTGCC	1641
Qy	366	TyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThrHisProGluSer	385
Db	1642	TACACCATCATCAACGGAAACCCCGGCAGAGCTTTGAAATCCACACCAACCCCTCAAAAC	1701
Qy	386	AsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAlaLysAsnGlnHisThr	405
Db	1702	AACGAAGGAGTCTTTCTGTGTTCAAACCATTTGACTATGAAATTTCTGCTTCTCCACACC	1761
Qy	406	LeuTyThrValGluValThrAsnGluAlaProPheValLeuLysLeu-----ProThr	422

RESULT 11

```

US-08-332-638-61
; Sequence 61, Application US/08332638
; Patent No. 5646250
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Warshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun

```


Qy 463 ThrAlaGluAspProAspLys---GluAsnGlnLysIleSerTyRArgIleLeuArgAsp 481
Db 1942 AATGCCAGGACCCGAGTCCCTGCAGCATCAACACCATCAGGTATCTGTTCACAGGAC 2001
Qy 482 ProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThrLeu 501
Db 2002 CCAGCAGGTTGGCTGAATATTAAACCCATCAACGGGACTGTTGACACACACGCTGTCTG 2061
Qy 502 AspArgGluAspGluGlnPheValArgAsnAsnIleTyRValMetValLeuAlaMet 521
Db 2062 GACCGTGAG---TCCCATTTGTGACACACAGCGGTACACTGCTCTCTTCCTGGCAATT 2118
Qy 522 AspAsnGlySerProProThrThrGlyThrLeuLeuLeuLeuLeuLeuLeuLeuVal 541
Db 2119 GACAGTGCAACCCCTCCGCTACGGGCACTGGGACTTTGTGATAAACCCCTGGAGGAGTG 2178
Qy 542 AsnAspHisGlyProValProGluProArgGlnIleThrLeuCys-----Asn 557
Db 2179 AATGACAAAGCCCGTTCATTATCCACAGTACAGTCTGATGATGTCGCAAAAC 2238
Qy 558 GlnSerProValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSer 577
Db 2239 CTCAGTGTAGTC-----ATTTGGGAGCATCAGATAAGGATCTTCACCCGATACAGAT 2292
Qy 578 ProPheGlnAlaGlnLeuThrAspAspSer-----AspIleTyRTrpThrAlaGluVal 595
Db 2293 CCTTCAAAATTGAAATCCCAACAAAGCTGTTCTGTATAAGTCTGG-----AAGATC 2346
Qy 596 AsnGluGluGlyAspThr---ValValLeuSerLeuLysLysPheLeuLysGlnAspThr 614
Db 2347 TCCAAGATCAACATACACACGCCCTGGTAGCTTCTTCAAAATCTGAACAAAGCAAC 2406
Qy 615 TyrAspValHisLeuSerLeuSerAspHisGlyAsn-----LysGluGlnLeuThrVal 632
Db 2407 TACAACCTGCCATCGTGTGACAGATTTCAGGGAACCCACCCATGACGAATATCACAGAT 2466
Qy 633 IleArgAlaThrValCysAspCysHisGlyHisValGluThrCysProGlyProTrpLys 652
Db 2467 CTCAGGTACAGTGTCTCTCGAGGAATTCGAAGTGGACTGCAACGGCGGGGGGCC 2526
Qy 653 GlyGlyPheLeuProValLeuGlyValAlaValLeuAlaLeuLeuPheLeuLeuVal 672
Db 2527 CTGGCCTTCAGCTGCC-----TCAGTCTGCTCTCTCAGCTCTTCAGCTTAGCTGT 2580
Qy 673 LeuLeuLeuLeuVal-----ArgLysLys 680
Db 2581 CTGTGAGACCTCTGAGCTGTGAAGCTTGACTCCCAAGTTTCCATAGCAACAGGAAAAA 2640
Qy 681 ArgLys 682
Db 2641 AAAAAA 2646

RESULT 12

US-08-474-067-3
; Sequence 3, Application US/08474067
; Patent No. 5811518
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/474,067
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/213,361
FILING DATE: 14-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/607,293
FILING DATE: 30-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1682
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2779 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 40..2191
US-08-474-067-3
Alignment Scores:
Pred. No.: 1,6e-90 Length: 2779
Score: 977.50 Matches: 265
Percent Similarity: 49.81% Conservative: 127
Best Local Similarity: 33.67% Mismatches: 309
Query Match: 22.37% Indels: 86
DB: 1 Gaps: 22
US-09-916-849A-1 (1-829) x US-08-474-067-3 (1-2779)
Qy 10 SerLeuLeuLeuGlnValCysTrpLeuGlnCysAlaAlaSerGluProCysArgAla 29
Db 67 TCTTCTCTCTGTCAGGCTTCTGTCTGTCGTCGTCAGAGATTAGATTGACCCCT 126
Qy 30 ValPheArgGluAlaGluValThrLeuGluAlaGlyGlyAlaGluGlnGluProGlyGln 49
Db 127 GGATTCACCAAAAGGTTTTTATATTGAACAGCCATTGATTTCACAGGACGACGCCA 186
Qy 50 AlaLeuGlyLysValPheMetGlyCysProGlyGlnGluProAlaLeuPheSerThrAsp 69
Db 187 ATTCTGAACCTGTGTGTGATGACTGCAAGGGGAAATAACAAATTGAACTTCGAAAGTTCT 246
Qy 70 AsnAspAspPheThrValArgAsnGlyGluThrValGlnGluArgArgSerLeuLysGlu 89
Db 247 AACCCAGACTTTAAGGTGGAACACGATGATCTTAGTTGCACTGAGNATGATCAGAA 306
Qy 89 ----- 89
Db 307 GCTGGCAGAGCTTTGTTGTCATGCAGCGTCTGAGCATGCTGAGGATATGCGAGAAAT 366
Qy 90 -----ArgAsnProLeuLys---IlePheProSerLysArg 100
Db 367 TTGATTGTTGGAGCTGATGAGAGGACGATGATTAAGGAAATCTTTAAGATAGAGGC 426
Qy 101 IleLeu-----ArgArgHisLysArgAspTrpValValAlaProIleSerValProGlu 118
Db 427 AACCTTGAATTCACAGACAAAAAGGGCTATTCTGGCGACTCCAATATTAATTCAGAA 486
Qy 119 AsnGlyLysGlyProPheProGlnArgLeuAsnGln---LeuLysSerAsnLysAspArg 137
Db 487 AATCAAGACCCACCATTTCCAGATCAGTTGCGAGGTCATCAGGAGTGAAGGACAGAG 546
Qy 138 AspThrLysIlePheTySerIleThrGlyProGlyValAlaAspSerProGluGlyVal 157
Db 547 GGAGCAAG-----TTCCGACTCTCTGTAAGGGAGTAGATCAAGACCCGAAAGGAAT 600

158 PheAlaValGluLysGluThrGlyTyrLeuLeuLeuAenLysProLeuAspArgGlu 177
QY : : : : :
Db TTTAGAATCAATGATCATGGCGGATGCTCTGTGAACCGACCCTGGATAGAGAACA 660

178 IleAlaIysTyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAsp 197
QY : : : : :
Db ATAGCCAATTATGACTGGAAAGTTGAAGTAACCGAATTTAAGTGGGAAAAATCATTGATGGC 720

198 ProMetAsnIleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGln 217
QY : : : : :
Db CCAGTCGCCTAGATAATTTCTGTTATTGATCAAATGATAACAGGCCGATGTTCCAAAGAA 780

218 AspThrPheArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnVal 237
QY : : : : :
Db GGACCCATTGTGTGTGTCACGTCATGGAAAGATCCCCTACAGGAACAACATGTGATCGGATG 840

238 ThrAlaThrAspGluAspAspAlaIleTyrThrTyAsnGlyValValAlaTyrSerIle 257
QY : : : : :
Db ACAGCAITTTGATGCTGATGATCCT--AGCACAGACAAGCTCTTCTCGGATATAACATC 897

258 HisSerGlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGly 277
QY : : : : :
Db CTCAGCAGACACCTACCAAACCTCCCAATAATGTTCTACATTGACCCAGAAAAGGGA 957

278 ThrIleSerValIleSerSerGly----LeuAspArgGlu-----LysValProGlu 293
QY : : : : :
Db GATATTGTCAGTGGTGTGCGCTGTACTGTGCTGATCTGTAGACAATGAAACCCGAG 1017

294 TyrThrLeuThrIleGlnAlaThrAspMetAspGlyAsp-----GlySerThrThr 311
QY : : : : :
Db TACGAGCTGGTTATTGAAGCAAGGATATGGCGGCCATGATGTGGGACTTACTGGAACT 1077

312 AlaValAlaValValGluIleLeuAspAlaAsnAspAsnAlaProMetPheAspProGln 331
QY : : : : :
Db GCRACTGCCATATCTTTATTGATGACAAAACGACCCACAGAAATTTACCAAGAAG 1137

332 LysTyrGluAlaHisValProGluAsnAlaValGlyHisGluValGlnArgLeuThrVal 351
QY : : : : :
Db GAGTTTCAGGCCACAGATAAGGAAGGAGTCACAGGA--GTAATAGTAAACTTAACGTGT 1194

352 ThrAspLeuaspAlaProAsnSerProAlaTyrArgAlaThrTyrLeuIleMetGlyGly 371
QY : : : : :
Db GGTACCCGAGATACCCAGCAACTGGAGCATGGAGAGCTGTCTACACTATTATTAAACGGA 1254

372 AspAspGlyAspHisPheThrIleThrHisProGluSerAsnGlnGlyIleLeuThr 391
QY : : : : :
Db AATCCAGGCCAGAGTTTGAATCCATACCAATCCCAGACTAATGAGGGAATGCTCTCT 1314

392 ThrArgLysGlyLeuAspPheGluAlaLysAsnGlnHisThrLeuTyrValGluValThr 411
QY : : : : :
Db GTTGCTAAACCTTTAGACTATGAGATTTCAGCATTTTCACACTGCTGATATAAAGTAGAA 1374

412 AsnGluAlaProPheValLeuLysLeu-----ProThrSerThrAlaThrIleVal 428
QY : : : : :
Db AATGAAGACCCGTGTGATTCAGACATAGCCTACGCTCCAGTTCACAGCAAGATTCAG 1434

429 ValHisValGluAspValAsnGluAlaProValPheValProProSerLysValValGlu 448
QY : : : : :
Db ATCCCGTTGAGGATGGAATGAAGCGCCTGTGTTTCCACCCAAACCCAAATCAGACTGACA 1494

449 ValGlnGluGlyIleProThrGlyGluProValCysValTyrThrAlaGluAspProAsp 468
QY : : : : :
Db AAAACAAGAGAATCCCTATTGGCAGCATTTGGTTAACAGTAATGCCACTGATCCAGAT 1554

469 Lys---GluAsnGlnLysIleSerTyrArgIleLeuArgAspProAlaGlyTyrLeuAla 487
QY : : : : :
Db ACTTTGCAACATCAGACTATCAGTATTTCAGTTTACAGGATTCAGCAAGCTGGCTAGAG 1614

488 MetAspProAspSerGlyGlnValThrAlaValGlyThrLeuAspArgGluAspGluGln 507
QY : : : : :
Db ATTAAATCCCCCAATGATACCGTTGGCCACCATGCTGTCTCGATTCGGGAATCTCCGCAT 1674

508 PheValArgAsnAnIleTyrGluValMetValLeuAlaMetAspAsnGlySerProPro 527
QY : : : : :

[illegible]

```

RESULT 13
US-08-474-068A-3
; Sequence 3, Application US/08474068A
; Patent No. 58157525
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,068A

```



```

; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1683
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2779 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 40..2191
; US-08-474-068A-3

Alignment Scores:
Pred. No.: 1 6e-90 Length: 2779
Score: 977.50 Matches: 265
Percent Similarity: 49.81% Conservative: 127
Best Local Similarity: 33.67% Mismatches: 309
Query Match: 22.37% Indels: 86
DB: 2 Gaps: 22

US-09-916-849A-1 (1-829) x US-08-474-068A-3 (1-2779)

QY 10 SerLeuLeuLeuGlnValCysTrpLeuGlnCysAlaAlaSerGluProCysArgAla 299
Db 67 TCCTTTCTGCTGCCAGGTTCTGTGTGTCGCTGCAGAAATTTAGAATGCACCCCT 126
QY 30 ValPheArgGluAlaGluValThrLeuGluAlaGlyGlyAlaGluGlnGluProGlyGln 49
Db 127 GGATTCACGAAAGGTTTTTATATTGAACAGCCATTGNAATTCACAGAGGACCGCCA 186
QY 50 AlaLeuGlyLyValPheMetGlyCysProGlyGlnGluProAlaLeuPheSerThrAsp 69
Db 187 ATTCTGAACCTGGTGTGTTGATGACTGCAAGGGGAATAACAAATTCGAAGTTTCT 246
QY 70 AsnAspPheThrValArgAsnGlyGluThrValGlnGluArgArgSerLeuLyGlu 89
Db 247 AACCCAGATTTAAGTGGAACACGATGGATCTTTAGTTGCACCTGAAGAAATGATCAGAA 306
QY 89 ----- 89
Db 307 GCTGGCAGAGCTTTGTTTGTCCATGCACGGTCTGAGCATGCTGAGGATATGGCAGAAAT 366
QY 90 -----ArgAsnProLeuLyS---llePheProSerLySArg 100
Db 367 TTGANTTTGGAGCTGATGAGAAGCACCATTCATTAAGGAAATCTTTAAGATAGAAGGC 426
QY 101 lleLeu-----ArgArgHisLySArgAspTrpValValAlaProIleSerValProGlu 118
Db 427 AACCTTGAATTCAGACAAAAGGCTATTCTGGCGACTCCAATATTATTCAGAA 486
QY 119 AsnGlyLyGlyProPheProGlnArgLeuAsnGln---LeuLySserAsnLySAspArg 137
Db 487 AATCAAAGACCAACCATTTCCAGATTCAGTTGGCAAGGTCATCAGGAGTGAAGGACAGAG 546
QY 138 AspThrLySilePheTyTrSerIleThrGlyProGlyAlaAspSerProProGluGlyVal 157
Db 547 GGAGCAAG-----TTCGACTCTCTGGTAAGGAGTAGATCAAGACCGGAAAGAAATT 600
QY 158 PheAlaValGluLySgluThrGlyTrpLeuLeuLeuAsnLySProLeuAspArgGluGlu 177

```


Db 601 TTTAGATCAATGAGATCAGTGGGATGTCTCTGTGACCCGACCCCTCGATAGAGACA 660
Qy 178 lleaLysTyrgluLeuPheClyHisAlaValSerGluAsnGlyAlaSerValGluAsp 197
Db 661 ATAGCCAAATATGAGCTGGAATGAAGTAAAGGATTTAAGTGGGAAATCAATGATGGC 720
Qy 198 ProMetAsnIleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGln 217
Db 721 CCAGTCCGCCATGATATTCTGTATTGATCAAAATGATACAGGCCGATGTTCAAGAA 780
Qy 218 AspThrPheArgGlySerValGluGlyValLeuProGlyThrSerValMetGlnVal 237
Db 781 GGACCTATGTTGGTCACTGATGAGGATCCCTACAGGAACAACCTGTGATGGGATG 840
Qy 238 ThrAlaThrAspGluAspAlaIleTyrrThrTyraAsnGlyValAlaIleSerIle 257
Db 841 ACAGATTGATGCTGATGATCTCT---AGCACAGAACGCTCTCTCGGTATTAACATC 897
Qy 258 HisSerGlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGly 277
Db 898 CTCAGCAGACACCTTACCACACCTTCCCAATATGTTCTACTTGCACCCAGAAAGGA 957
Qy 278 ThrIleSerValIleSerSerGly-----LeuAspArgGlu-----LysValProGlu 293
Db 958 GATATTGTCAGTGGTGTCCCTGTACTGTCTGTGATGTGAGACAATGGAAACCCGGAAG 1017
Qy 294 TyrThrLeuThrIleGlnAlaThrAspMetAspGlyAsp-----GlySerThrThrThr 311
Db 1018 TACGAGCTGTTATTGAAGCCAGGATATGGCGCCATGATGTGGACTTACTTGGAACT 1077
Qy 312 AlaValAlaValValGluIleLeuAspAlaAsnAspAlaProMetPheAspProGln 331
Db 1078 GCAACTGCCATATTCTTATTGATGACAAACACCCACCCACCAAGAAATTTACCAAGAG 1137
Qy 332 LysTyrgluAlaHisValProGluAsnAlaValGlyHisGluValGlnArgLeuThrVal 351
Db 1138 GAGTTTCAGGCCACAGTAAGAGAGGTACACAGA---GTAATGATAAATTAACITGTT 1194
Qy 352 ThrAspLeuAspAlaProAsnSerProAlaTrpArgAlaThrTyrrLeuIleMetGlyGly 371
Db 1195 GGTGACCGAGATGACCCAGCAACTGGAGCATGGAGAGCTGTCTACATATTATTAAACGA 1254
Qy 372 AspAspGlyAspHisPheThrIleThrHisProGluSerAsnGlnGlyIleLeuThr 391
Db 1255 AATCAGGCGAGATTTGAAATCCATACCNAATCCCGAGCTAATGAGGAATGCTCTCT 1314
Qy 392 ThrArgLysGlyLeuAspPheGluAlaLysAsnGlnHisThrLeuTyrrValGluValThr 411
Db 1315 GTTGTCAAACTTTAGACTATGAGATTTACAGATTTTACACATTTCTGATAAAGTAGAA 1374
Qy 412 AsnGluAlaProPheValLeuLysLeu-----ProThrSerThrAlaThrIleVal 428
Db 1375 AATGAAGACCCGTTGATTCAGACATAGCCCTACGGTCCCGAGTTCCACAGCAACAGTTTCA 1434
Qy 429 ValHisValGluAspValAsnGluAlaProValPheValProSerLysValValGlu 448
Db 1435 ATCACCCTTGAGATGATGAATGAGGCCCTGTTTCCACCCAAACCCAAATGACAGTGACA 1494
Qy 449 ValGlnGluGlyIleProThrGlyGluProValCysValTyrrThrAlaGluAspProAsp 468
Db 1495 AAACAAGAGAACATCCCTTATTTGGCAGCATTTGTGTTAAAGTAAATGACCTGATCCAGAT 1554
Qy 469 Lys---GluAsnGlnLysIleSerTyrrArgIleLeuArgAspProAlaGlyTrpLeuAla 487
Db 1555 ACITTTGCAACATCAGACTATCAGTATTTCAGTTTACAGGATCCAGCAAGCTGGCTAGAG 1614
Qy 488 MetAspProAspSerGlyGlnValThrAlaValGlyThrLeuAspArgGluAspGluGln 507
Db 1615 ATTAATCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1674
Qy 508 PheValArgAsnIleTyrrGluValMetValLeuAlaMetAspAsnGlySerProPro 527
Db 1675 ---GTTCAGGATAACAATACACTGTCTCTTCTGCAATACAGATGATGATGATGATGATGAT 1731

Qy 528 ThrThrGlyThrGlyThrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 547
Db 1732 GCTACAGGTACAGAACTTTTACACATCCTTGGAGGAGCTCAATGACAAATGTCCTCC 1791
Qy 548 ProGluProArgGlnIleThrIleCysAsnGlnSerProValArgHis-----ValLeu 565
Db 1792 CTTTACCACACACTGGCAAAAGTCTGTGATGATGATGATGATGATGATGATGATGATGAT 1851
Qy 566 AsnIleThrAspLysAspLeuSerProHisThrSerProPheGlnAlaGlnLeuThrAsp 585
Db 1852 GGAGCATCAGACAAAGACCTCCATCCCAACACAGATCCATTTAAATTTGAACCTGAGTAAG 1911
Qy 586 AspSerAspIleTyrrThrAlaGluValAsnGluGluGlyAspThr---ValValLeu 604
Db 1912 CAATCTGGTCCAGAAAGTATGGAGATCAACAGCTTAACAATACTCATGCCAGGTT 1971
Qy 605 SerLeuLysPheLeuLysGlnAspThrTyrrAspValHisLeuSerLeuSerAspHis 624
Db 1972 GTCCTGCTTCAAAACCTGAAAGGCCAATTACACATCCCAATCTCAGTACAGATTTCT 2031
Qy 625 GlyAsnLysGluGlnLeu-----ThrValIleArgAlaThrValCysAspCys-Hi 641
Db 2032 GGA---AAACCACTCTGACTAACACACAGAACTGAAATTTACAGTGTCTTCTGCAAG 2088
Qy 641 sGlyHisValGluThrCysProGlyProTyrrLysGlyGlyPheIleLeuProValLeucl 661
Db 2089 AAT-----CCAGATGGACTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGT 2130
Qy 661 yAlaValLeu-----AlaLeuLeuPheLeuLeu----- 670
Db 2131 AGCATGACTTATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2190
Qy 671 -----LeuValLeuLeuLeuValArgLysLysArgLysLysLysLysLysLysLys 686
Db 2191 TAAGCATTAAGAGTATTATTTATCTGCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCT 2248
Qy 686 oLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyrrTyrrGlyGlyGlyGly 706
Db 2249 -----AGAGTTAACTTTATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2283
Qy 706 yGlyGluGluAspGlnAspTyrrAspIleThrGlnLeuHisArgGlyLeuGluAlaArgPr 726
Db 2284 TATACAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2341
Qy 726 oGluValValLeuArgAsn 732
Db 2342 -GTTGTCAATTAAGAAAT 2359

RESULT 15

US-09-016-434-1125
; Sequence 1125, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016.434

```

; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: 2A-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1125:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2880 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g1389852
;
US-09-016-434-1125

Alignment Scores:
Pred. No.: 2,158-90 Length: 2880
Score: 976.50 Matches: 266
Percent Similarity: 50.51% Conservative: 127
Best Local Similarity: 24.19% Mismatches: 322
Query Match: 22.35% Indels: 53
DB: 4 Gaps: 25

US-09-916-849A-1 (1-829) x US-09-016-434-1125 (1-2880)
QY 90 ArgAsnProLeuLys---IlePheProSerLysArgIleLeuArgA:GHISLysArgAsp 108
DB 416 AGAAACCAACCAACACATTTGAAGTGAACCAAGTCCATCATCTGCCAAGGGGA 475
QY 109 TrpValValAlaProIleSerValProGluAsnGlyLysGlyProPheProGluArgLeu 128
DB 476 TGGGTATGGAATCACTCTTTGTTTGAAGAAACATATGGACCAAGATCTCTCAGATGTT 535
QY 129 AsnGlnLeuLysSerAsnLysAspArg---AspThrLysIlePheTyrSerIleThrGly 147
DB 536 GGAAGAGTGCATCTCAATTCGACAAAGGTGATGATCTGCAAGTACATCTTACTTGA 595
QY 148 ProGlyValAspSerProProGluGlyValPheAlaValGluLysGluThrGlyTrpLeu 167
DB 596 GAGGTGTGGGACT-----ATATTATCATTCAGCATACCAACGGGTGATATC 643
QY 168 LeuLeuAsnLysProLeuAspArgGluGluLeuAlaLysTyrGluLeuPheGlyHisAla 187
DB 644 CACTCAACAAAAGCCTAGACAGAGAGAGAACCCACTATGCTTCATGCTCAAGCT 703
QY 188 ValSerGluAsnGlyAlaSerValGluAspProMetAsnIleSerIleIle---ValThr 206
DB 704 ATTGATAGCGTACAAACAACTCTTGAGCTGTAATCCGAGTTTCATCATCAAGTGCAA 763
QY 207 AspGlnAsnAspHisLysProLysPheThrGlnAspThrPheArgGlySerValLeuGlu 226
DB 764 GACATCAATGACACCGCTCCAAATTCACAGATGACCATACATTTGTTACTGTGCTGAA 823
QY 227 GlyValLeuProGlyThrSerValMetGlnValThrAlaThrAspGluAspAlaIle 246
DB 824 ATGTGATATGGTACCTCTCTCTTCTACAGTGACAGCTACTGATGACAGTACCTTACC 883
QY 247 TyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGlnGluProLysAspProHis 266
DB 884 TATGGAACAGCGCTCGGTGGTGTACAGCATCTCCAGGACACCC-----931
QY 267 AspLeuMetPheThrIleHisArgSerThrGlyThrIleSerValIleSerSerGlyLeu 286
DB 932 -----TACTTCTCGCTGACCCCTAAACAGAGGATTTATAGACCGCCTTACATACATG 985

```

```

287 AspArgGluLysValProGluTyrThrLeuThrIleGlnAlaThrAspMetAspGly--- 305
986 GACAGAGAACCCAGAGAACATTTACTCCGTAGCTATTCAAGCCAAAGACATGGCTGGGCAA 1045
306 -----AspGlySerThrThrAlaValAlaValGluIleLeuAspAla 321
1046 GTTGAGGGCTTTTCAGGATCTACACAGTCACACATCACC-----TTAACCGATGTC 1096
322 AsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsnAla 341
1097 AATGACACCCACACACGCTTTCTCTCAAAAACATATCAGCTATATGTTCTCTGAGTCAGCT 1156
342 ---ValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
1157 CAAAGTTGGTTGAGCTGTTGGGAAATCAAGGCAATGATGCTGACACTGGCTCAATGCT 1216
361 AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
1217 -----GACATGACCTACTCCATCATATAATGGTATGCGATGGGAATATTTCTCAATCTCC 1270
381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
1271 ACTGACAAAGACAGACAGAGAGGAATCCTTTTAAAGAACCCACTGAATCTAGAGAA 1330
401 LysAsnGlnHisThrLeuTyrValGluValThrAsn-----GluAlaProPheVal 417
1331 AAGAAGTCATATACCTTCAACATAGAGAGCAATATACATCTTGATTTTCGCTTTCT 1390
418 LeuLysLeuProThrSerThrAlaThrIleVal---ValHisValGluAspValAsnGlu 436
1391 CACTTGGGTCCTTTTAAAGATGCTACTATGCTGAAGATCAATTTGTTGGGATGTAGATGAA 1450
437 AlaProValPheValProProSerLysValValGluValGlnGluGlyIleProThrGly 456
1451 CCACCATATTTTCCATGCTCTTCTACCTCATGGAAGTCTACGAAATGCCAAGATTGGG 1510
457 GluProValCysValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyr 476
1511 ACCGTCGTTGGTACAGTTTGGCACACAGATCTCGACAGTACTAACAGCTTAGTAGATAC 1570
477 ArgIle-----LeuArgAspProAlaGlyTrpLeuAlaMetAspProAspSerGly 493
1571 TTCATCACTACATGTTTGAAGACGACAGA---TTTTTCAACATTTGATGCCCAATCTGGG 1627
494 GlnValThrAlaValGlyThrLeuAspArgGluAspGluGlnPheValArgAsnAsnIle 513
1628 ACCATTAGACTACAAAGGTTCTGCACAGAGAGAAACTCCATGG-----1672
514 TyrGluValMetValLeuAlaMetAspAsnGlySerProProThrThrGlyThrGlyThr 533
1673 TACAACATCACAGTCATCTGCTTCAAGAAATTTGATAATCTCTGATTTGCTGAGCCATGTCAC 1732
534 LeuLeuLeuThrLeuLeuAspValAsnAspHisGlyPro---ValProGluProArgGln 552
1733 GTGGTATTAGAGTTCTGGATGTCATGACATTCACCCGAACTTCCAGGGAATATGAT 1792
553 IleThrIleCysAsnGlnSer-----ProValArgHisValLeuAsnIleThrAsp 569
1793 ATTATTGATGTGAAATTTCTAAGCCTGGCCAGGTTATTTCATACCATCATGCTCCACTGAT 1852
570 LysAspLeuSerProHisThrSerProPheGlnAlaGlnLeuThrAspAspSerAspIle 589
1853 AAAGATGATTTTCCCAATGACCAAGGTTTAACTCTTCTTGTGATGAACCCCTGCTGTGA 1912
590 -----TyrTrpThrAlaGluValAsnGluGlyIleAspThrValValLeuSerLeuLys 607
1913 AATCCAACTTCACTCTGAGGACATGAAGATACACAGCCAGCATCTCTGCAAGGCGG 1972
608 LysPheLeuLys-----GlnAspThrTyrAspValHisLeuSerLeuSerAspHis 624
1973 AGGAGATTTAGTCGAATCTGTTCCAGGATGTATTATCTGCCCATTTATGATCTCTGATGT 2032

```

Search completed: September 23, 2004, 01:28:12
Job time : 255 secs

This Page Blank (uspro)

421 ProThrSerThrAlaThrIleValValHisValGluuValAspValAsnGluuAlaProValPhe 440
 Db CCAACCTCCACAGCCACCATAGTGGTCCAGTGGAGATGTGATGAGGACCTGTGT 1373
 441 ValProSerTyrValValGluuValGluuGlyIleProThrGlyGluuProValCys 460
 Db GTCCACCCCTCCAAAGTCGTAGGTCAGAGGGGATCCCACTGGGAGCGCTGTGT 1433
 461 ValTyrThrAlaGluuAspProAspLysGluuAsnLysIleSerTyrArgIleLeuAig 480
 Db GTCTACACTGCAGAGACCCCTGCAAGAGAGATCAAAAGATCAGTACCGCATCTTGAGA 1493
 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
 Db GACCCAGCAGGTCGTAGCATGATGACCCAGACAGTGGCAGGTCAAGCTGTGGGCACC 1553
 501 LeuAspArgGluuAspGluuGlnPheValArgAsnAsnIleTyrGluuValMetValLeuAla 520
 Db CTCGACCGTCAGGATGAGCATTTGTGAGGAACAACATCTATGAAGTCATGGTCTTGCC 1613
 521 MetAspAsnGlySerProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
 Db ATGACATATGAGACCCCTCCACCACTGGCAGGAAACCTTCTGCTAACACTGATTGAT 1673
 541 ValAsnAspHisGlyProValProGluuProArgGlnIleThrIleCysAsnGlnSerPro 560
 Db GTCAACGACCATGTCGCCAGTCCCTGAGCCCGTCAGATCAACATCTGCAACCAAGGCCCT 1733
 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
 Db GTGGCCACAGTGTGAACATCAGGACAGGACCTGTCTCCACACACTCCCTTCCAG 1793
 581 AlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluuValAsnGluuGlyAsp 600
 Db GCCCAGCTCACAGATGACTCAGACATCTACTGGAGCGCAGAGGTCAACAGGAGGTGAC 1853
 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
 Db ACAGTGTCTTGTCTCCCTGAAGAGTCTCTGAAGCAGGATCATATGACGTGACCTTCT 1913
 621 LeuSerAspHisGlyAsnLysGluuLeuThrValIleArgAlaThrValCysAspCys 640
 Db CTGTCTGACCATGCAACAAAGAGCAGTGCAGGTGATGAGGCACTGTGTGGAGCTGC 1973
 641 HisGlyHisValGluuThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu 660
 Db CATGCCCATGTCCAAACCTGCCCTGGACCCCTGGAAAGAGGTTTCATCTCCCTCTGCTG 2033
 661 GlyAlaValLeuAlaLeuPheLeuLeuValLeuLeuValLeuValArgLysLys 680
 Db GGGCTGTCCTGCT 2093
 681 ArgLysIleLysGluuProLeuLeuLeuProGluuAspAspThrArgAspAsnValPheTyr 700
 Db CGGAAGATCAAGAGCCCTCTACTCCAGAAAGATGACCCCGTCAAGCTCTCTCTCTCTCT 2153
 701 TyrGlyGluuGlyGlyGlyGluuAspGlnAspTyrAspIleThrGlnLeuHisArg 720
 Db TATGGCAAGAGGGGGTGGGAGAGGACCAAGGACTATGACATCAACCCAGCTCCACCGA 2213
 721 GlyLeuGluuAlaArgProGluuValValLeuArgAsnAspValAlaProThrIleIlePro 740
 Db GGTCTGGAGGCCAGCCGAGGTGTTCTCCGAATGACGTGGCAGCAACCAACCATCATCCG 2273
 741 ThrProMetTyrArgProArgProAlaAsnProAspGluuIleGlyAsnPheIleIleGlu 760
 Db ACACCATGTACCGTCTAGGCGACCAACCCAGATGAAATCGGCACTTTATTAATTTGAG 2333
 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
 Db AACCTGAGGCGGCTAACACAGACACCCAGCCCGCTTACGACACCCCTCTCTGTGTCTC 2393

781 AspTyrGluuGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
 Db GACTATGAGGCGAGCGGCTCCGAGCGCGTCCCTGAGCTCCCTCAGCTCCCTCCGCTCC 2453
 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluuTyrPheLysLysLeuAla 820
 Db GACCAAGACCAAGATGATGATTAATCTGACAGTGGGCGAGCCCTTCAAGAGCTGGCA 2513
 821 AspMetTyrGlyGlyGlyGluuAsp 829
 Db GACATGTACGCTGGCGGAGGAGCAGC 2540
 RESULT 3
 US-09-880-107-3759
 ; Sequence 3759, Application US/09880107
 ; Patent No. US20020142981A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Horne, Darci T.
 ; APPLICANT: Vockley, Joseph G.
 ; APPLICANT: Scherf, Uwe
 ; APPLICANT: Gene Logic, Inc.
 ; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
 ; FILE REFERENCE: 44921-5028-WO
 ; CURRENT APPLICATION NUMBER: US/09/880,107
 ; CURRENT FILING DATE: 2001-06-14
 ; PRIOR APPLICATION NUMBER: US 60/211,379
 ; PRIOR FILING DATE: 2000-06-14
 ; PRIOR APPLICATION NUMBER: US 60/237,054
 ; PRIOR FILING DATE: 2000-10-02
 ; NUMBER OF SEQ ID NOS: 3950
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 3759
 ; LENGTH: 3171
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: Genbank Accession No. US20020142981A1 X63629
 US-09-880-107-3759
 Alignment Scores:
 Pred. No.: 0 Length: 3171
 Score: 4369.00 Matches: 829
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0
 US-09-916-849a-1 (1-829) x US-09-880-107-3759 (1-3171)
 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln 20
 Db ATGGGGCTCCCTCGTGGACCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113
 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
 Db TGGCGGCTCCGAGCCGTCGCGGGGTCTTCAGGAGGCTGAAGTCACTTGGAGCG 173
 41 GlyGlyValaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
 Db GGAGGCGGAGCAGGAGCCCGCCAGGCGCTGGGGAAGTATTCATGGCTGGCTGG 233
 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
 Db CAAGACCCAGCTCTGTTTACGACTGATATGATGATGATGATGATGATGATGATGATGAT 293
 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
 Db GTCCAGAAAGAGGTCACTGAAGGAAGAAATCCATTGAAGATCTCCCATCCAAAGCT 353
 101 IleLeuArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
 Db ATCTTACGAGACACAAAGAGATTGGGTGGTTCCTCCATATCTCTCTCTCTCTCTCTCTCT 413

Qy	121	LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys	141
Db	414	AAGGGTCCTCCCTCCCCACAGAGACTGAATCAGCTCAAGTCTAATAAAGATAGAGACACCAAG	473
Qy	141	IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal	160
Db	474	ATTTCTACAGCATCACGGGCGGGGCGAGACGCCCTGAGGGTGTTCTTCGCTGTA	533
Qy	161	GluLysGluThrGlyTyrTrpLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys	180
Db	534	CAGAAGGAGACAGGCTGTTGTTGTAATAAGCCACTGGACCGGAGGAGATTGCCAAG	593
Qy	181	TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetaAsn	200
Db	594	TATGAGCTCTTTGGCCACGCTGTGTACAGAAATGGTGCTCAGTGGAGGACCCCATGAAC	653
Qy	201	IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe	220
Db	654	ATCTCCATCATCGTGACCGACCAAGATGACCAAGCCCAAGTTTACCCAGACACCTTC	713
Qy	221	ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr	240
Db	714	CGAGGGAGTGTCTTAGAGGGAGTCTCTACAGGTACTTCTGTGATCGAGGTGACAGCCACA	773
Qy	241	AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln	260
Db	774	GATCAGGATGATGCCACTACACTACATGGGTGGTGTGTTACTTCATCCATGAGCAA	833
Qy	261	GluProLysAspProHisAspLeuValPheThrIleHisArgSerThrGlyThrIleSer	280
Db	834	GAACCAAGGACCCACACAGCTCATGTGTCAATTACCCGAGACAGAGCACCATCAGC	893
Qy	281	ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla	300
Db	894	GTCAATCCAGTGGCTTGGACCGGGAAAAAGTCCCTGAGTACACATGACCATCCAGGCC	953
Qy	301	ThrAspValAspGlyAspGlySerThrThrThrAlaValAlaValValGluIleLeuAsp	320
Db	954	ACAGACATGAGTGGGACGGCTCCACACACCGCAGTGGCAGTAGTGGAGATCCTTGAT	1013
Qy	321	AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn	340
Db	1014	GCCAAATGCAATGTCTCCATGTTTGACCCCGAAGTACGAGGGCCCATGTGCCCTGAGAAT	1073
Qy	341	AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro	360
Db	1074	GCAGTGGGCCATGAGGTGCAGAGCTGACGGTCACTGATCTGGACGCCCCCACTCACCA	1133
Qy	361	AlaTrpArgAlaThrTyrLeuIleMetGlyClyAspAspGlyAspHisPheThrIleThr	380
Db	1134	CGGTGGCGTGCCACTACCTTATCATGGCGCGTACGACGGGGACCAATTTTACCATCACC	1193
Qy	381	ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla	400
Db	1194	ACCCACCTTGAGACCAACCGGCATCTGTACACACCGAGGAGGTTGGATTTTGAGGCC	1253
Qy	401	LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu	420
Db	1254	AAAAACAGACACACCTGTACGTTGAGTGTACCAACAGGGGCCCTTTTGTGCTGAAGTC	1313
Qy	421	ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe	440
Db	1314	CCAACTTCCACAGCACCATAGTGGTCCAGTGGAGGATGTGATGAGGCACCTGTGTTT	1373
Qy	441	ValProProSerLysValValGluValGlnGlyIleProThrGlyGluProValCys	460
Db	1374	GTCCACCTCCAAAGTCGTTGAGTCCAGGAGGCATCCCCCATCTGGGAGCCTGTGTGT	1433
Qy	461	ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg	480
Db	1434	GTCTACACTGCAGAAGACCCCTGACAAGAGAAATCAAAAGATCAGCTACCGCATCCTGAGA	1493
Qy	481	AspProAlaGlyTyrTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr	500

RESULT 4

RESUL 4
US-10-058-270A-125

1494	Db	GA	CC	AG	CA	GG	GT	GG	GT	AG	CT	CA	TG	GA	CC	CA	AG	CA	GT	GG	CA	GG	GT	CA	CA	GT	GT	GG	CA	CC	1553
501	Qy	Leu	Asn	Arg	Glu	Asp	Glu	Gln	Phe	Val	Arg	Asn	Asn	Ile	Tyr	Glu	Val	Met	Val	Leu	Ala	520									
1554	Db	CT	CG	CC	GT	GA	GC	AT	TG	TG	TG	TG	GA	GA	CA	CA	AT	CT	AT	GA	AG	TC	AT	GA	GT	CT	GT	GG	CC	1613	
521	Qy	Met	Asp	Asn	Gly	Ser	Pro	Pro	Thr	Thr	Gly	Thr	Gly	Thr	Leu	Leu	Leu	Thr	Leu	Leu	Asp	540									
1614	Db	AT	GG	CA	AA	TG	CA	AG	CC	CT	CC	CA	CA	TG	GG	CA	CG	GA	AC	CC	CT	CT	CG	TAA	CAC	TG	AT	TG	AT	1673	
541	Qy	Val	Asn	Asp	His	Gly	Pro	Val	Pro	Glu	Pro	Arg	Gln	Ile	Thr	Ile	Cys	Asn	Gln	Ser	Pro	560									
1674	Db	GT	CA	CG	AC	CA	TG	CC	GT	CC	CT	TG	AG	CC	CC	GT	CAG	AT	CA	CA	AT	CT	GA	CA	CC	AA	AG	CC	CT	1733	
561	Qy	Val	Arg	His	Val	Leu	Asn	Ile	Thr	Asp	Lys	Asp	Leu	Ser	Pro	His	Thr	Ser	Pro	Phe	Gln	580									
1734	Db	GT	CG	CC	CA	CG	GT	GT	GA	CA	TC	AC	GG	CA	AG	CA	CT	GT	CT	CC	CA	CA	CC	CT	CC	CC	CT	TT	CC	AG	1793
581	Qy	Ala	Cln	Leu	Thr	Asp	Asp	Ser	Asp	Ile	Tyr	Trp	Thr	Ala	Scl	Val	Asn	Glu	Gly	Asp	600										
1794	Db	GCC	AG	CT	CA	CA	TG	ACT	CA	GA	CA	CT	ACT	TG	CA	GG	CA	GG	AG	GT	CA	AG	AG	AG	GT	GA	C	1853			
601	Qy	Thr	Val	Val	Leu	Ser	Leu	Lys	Lys	Phe	Leu	Lys	Gln	Asp	Thr	Tyr	Asp	Val	His	Leu	Ser	620									
1854	Db	AC	AG	TG	CT	TG	CT	CT	GA	GA	AG	TTC	TC	GA	AG	CA	GA	TAC	AT	CA	TG	CA	CG	TG	CA	CC	TT	CT	1913		
621	Qy	Leu	Ser	Asp	His	Gly	Asn	Lys	Glu	Gln	Leu	Thr	Val	Ile	Arg	Ala	Thr	Val	Cys	Asp	Cys	640									
1914	Db	CT	GT	CT	CA	CC	AT	GG	CA	CA	AA	AG	CA	GT	CA	GG	TG	AT	CA	GG	CC	CA	CT	GT	GT	CG	CA	CT	1973		
641	Qy	His	Gly	His	Val	Glu	Thr	Cys	Pro	Gly	Pro	Trp	Lys	Gly	Gly	Phe	Ile	Leu	Pro	Val	Leu	660									
1974	Db	CAT	GG	CA	AT	GT	CA	AA	AC	CT	GC	CT	GG	CA	CC	CT	GA	AA	AG	AG	TTC	AT	CT	CT	CC	CT	GT	CT	2033		
661	Qy	Gly	Val	Val	Leu	Ala	Leu	Phe	Leu	Leu	Leu	Val	Leu	Leu	Leu	Leu	Val	Arg	Lys	Lys	680										
2034	Db	GG	GG	CT	GT	CC	GT	CT	GT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	2093									
681	Qy	Arg	Lys	Ile	Lys	Glu	Pro	Leu	Leu	Pro	Glu	Asp	Asp	Thr	Arg	Asp	Asn	Val	Phe	Tyr	700										
2094	Db	CG	GA	AT	CA	AG	GAG	CC	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	CT	2153										
701	Qy	Tyr	Gly	Glu	Glu	Gly	Gly	Glu	Glu	Asp	Gln	Asp	Tyr	Asp	Ile	Thr	Gln	Leu	His	Arg	720										
2154	Db	TAT	TG	CG	CA	AG	GG	GG	GT	GG	CA	AG	AG	CA	GG	ACT	AT	GA	CA	TC	CA	CC	CA	CG	CT	CC	CA	CG	2213		
721	Qy	Gly	Leu	Glu	Ala	Arg	Pro	Glu	Val	Val	Leu	Arg	Asn	Asp	Val	Ala	Pro	Thr	Ile	Leu	Pro	740									
2214	Db	GG	CT	GG	AG	CC	AG	CG	GG	AG	TG	GT	TCT	CG	CA	AT	GA	CG	TG	GC	CA	CC	AT	CA	CT	CC	CG	2273			
741	Qy	Thr	Pro	Met	Tyr	Arg	Pro	Arg	Pro	Ala	Asn	Pro	Asp	Glu	Ile	Gly	Asn	Phe	Ile	Glu	760										
2274	Db	AC	ACC	AT	G	TAC	CG	CT	CT	TAG	CC	GAG	CA	CC	CA	CC	CAG	AT	GAA	AT	CG	CA	ACT	TTA	TAA	TG	AG	2333			
761	Qy	Asn	Leu	Lys	Ala	Ala	Asn	Thr	Asp	Pro	Thr	Ala	Pro	Tyr	Asp	Thr	Leu	Leu	Val	Phe	780										
2334	Db	AAC	CT	GA	GG	GG	CT	AAC	CA	GAG	CCCC	CA	CAG	CCCC	CC	CT	CAG	CA	CC	CT	CT	TGG	TG	TC	2393						
781	Qy	Asp	Tyr	Glu	Gly	Ser	Asp	Ala	Ala	Ser	Leu	Ser	Ser	Ser	Leu	Thr	Ser	Ser	Ala	Ser	800										
2394	Db	GAT	CT	AT	G	GG	CA	CG	CG	CT	CC	CA	GG	CG	GG	TC	CT	GAG	CT	CC	CT	CA	CC	CT	CC	CG	CT	CC	2453		
801	Qy	Asp	Gln	Asp	Gln	Asp	Tyr	Asp	Tyr	Leu	Asn	Glu	Trp	Gly	Ser	Arg	Phe	Lys	Lys	Leu	Ala	820									
2454	Db	GAC	CA	AG	CA	CA	GAT	TAC	GAT	TAT	CT	TA	CA	AG	AG	TG	GG	CA	GC	GC	TT	CA	AG	AG	CT	GG	CA	2513			
821	Qy	Asp	Met	Tyr	Gly	Gly	Glu	Asp	Asp	829																					
2514	Db	GAC	AT	G	TAC	GT	GG	CG	GG	AG	CA	CC	2540																		

RESULT 4

RESOL 4
US-10-058-270A-125

654 ATCTCCATCATCGTACCGACCGAGATGATCCACAAAGCCCAAGTTTACCCAGGACACCTTC 713
QY
221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
DB
714 CGAGGGAGTGTCTAGAGGGAGTCTTACCAGGTACTTCTGTGATGACAGGTGACAGCCACA 773
QY
241 AspGluAspAspAlaIleThrThrTyAsnGlyValValAlaValSerIleHisSerGln 260
DB
774 GATGAGGATGATGCATCTACATCACTACATGAGGTGGTGTCTTACTCATCCATAGCCAA 833
QY
261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
DB
834 GAACCAAGAGGCCACACAGCTCATGTTCACAAATCACCGGAGCAGACGACCATCAGC 893
QY
281 ValIleSerSerGlyLeuAspArgGluLysValProGluThrThrLeuThrIleGlnAla 300
DB
894 GTCACTCCAGTGGCTGGACCGGGAAAAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
QY
301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValGluIleLeuAsp 320
DB
954 ACAGACATGATGGGACCGCTCCACCAACGACGAGTGGCAGTAGTGGAGATCCTTGAT 1013
QY
321 AlaAsnAspAlaProMetPheAspProGlnLysTyrgLualHisValProGluAsn 340
DB
1014 GCCAATGCAATGCTCCCATGTTTACCCCCAGAAAGTACAGGGCCCATGTGCTGAGAT 1073
QY
341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
DB
1074 GCAGTGGGCCATGAGTGCAGAGCTGACGGTCACTGATCTGGACGCCCCCACTCACCA 1133
QY
361 AlaTrpArgAlaThrTyLeuIleMetGlyCysAspAspGlyAspHisPheThrIleThr 380
DB
1134 CGGTGGCGTGCACCTACCTATCATGGCGGTGACGACGGGACCAATTTACCATCACC 1193
QY
381 ThrHisProGluSerAsnGlnGlyLeuThrThrArgLysGlyLeuAspPheGluAla 400
DB
1194 ACCACCTTGAGACCAACCGGCATCTGCACACAGGAGGGTTTGGATTTTGGAGCC 1253
QY
401 LysAsnGlnHisThrLeuTyValGluValThrAsnGluAlaProPheValLeuLysLeu 420
DB
1254 AAAAACACAGCACACCTGTAGCTTGAAGTGCACCAACGAGGCCCTTTTGTCTGGAAGCTC 1313
QY
421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
DB
1314 CCACTTCCACAGCACCATAGTGTCCAGTGGAGGATGTGATGAGGACACCTGTGTTT 1373
QY
441 ValProSerLysValValGluValGlnGluGlyLeuProThrGlyGluProValCys 460
DB
1374 GTCCACCCCTCCAAAGTGGTGGGTCCAGAGGGGCATCCCACTGGGGAGCGCTGTGTGT 1433
QY
461 ValTyThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyArgIleLeuArg 480
DB
1434 GTCTACACTGCAGAGACCCCTGACAGAGAGAAATCAAAGATCAGCTACCCGATCCTGAGA 1493
QY
481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
DB
1494 GACCCAGCAGGCTGGCTAGCATCGACCCACAGAGTGGCAGGTCACAGTGTGGGCACC 1553
QY
501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrgLualMetValLeuAla 520
DB
1554 CTCGACCGTGGAGTGAAGATTTGTGTAGGAAACAAATCTATGAAGTCATGGTCTGGCC 1613
QY
521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
DB
1614 ATGGACATGGAGCCCTCCACCATCGGCACGGGAACCCCTTCTGCTTACACATGATTGAT 1673
QY
541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
DB
1674 GTCAACGACCATGGCCCACTGCTGAGCCCGTCCAGATCAGATCTCCAAACCAAGCCCT 1733
QY
561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
DB
1734 GTGGCCACGTGCTGAACATCAGGACAAAGACCTGTCTCCCCACACCTCCCTTTCAG 1793

QY 581 AlaGlnLeuThrAspAspSerAspIleTyTrpThrAlaGluValAsnGluGluGlyAsp 600
DB 1794 GCCCAGCTCACAGATGACTCAGACATCTACTGGCGCAGAGGTCAACGAGGAAGGTGAC 1853
QY 601 ThrValValLeuSerLeuLysIlyshPheLeuLysGlnAspThrTyArgValHisLeuSer 620
DB 1854 ACAGTGGTCTTGTCTCCCTGAAGAAGTTCCTGAAGCAGGATCATATGACGTGCACCTTCT 1913
QY 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
DB 1914 CTGTCTGACCATGGCAACAAGAGCAGCTGACGTGATCAGGGCCACTGTGTGGACTGC 1973
QY 641 HisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValLeu 660
DB 1974 CATGGCCATGTGAAACCTGCGTGCACCTGGAAGAGGAGGTTTTCATCTCCTCCTGTGCTG 2033
QY 661 GlyAlaValLeuAlaLeuPheLeuLeuValLeuLeuLeuLeuValArgLysLys 680
DB 2034 GGGGCTGTCTGGCTCTGTCT 2093
QY 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTy 700
DB 2094 CGGAAGATCAAGGACCCCTCTACTCTCCAGAAGATGACACCCGTGACACGCTCTTCTAC 2153
QY 701 TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyArgPheIleThrGlnLeuHisArg 720
DB 2154 TATGCGAAGAGGGGGTGGCAAGAGGACGAGGACTATGACATCACCAGCTCCACCGA 2213
QY 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
DB 2214 GTCTGGAGGCCAGCGCGAGGTGGTCTCCGAATGACGTGGCACCACCAACCATCCCG 2273
QY 741 ThrProMetTyArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
DB 2274 ACACCCATGTACCGTCTAGGCGACCCAGCAACCCAGATGAAATCGGCAACTTATTAATTGAG 2333
QY 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProTyArgPheThrLeuLeuValPhe 780
DB 2334 AACCTGAGGCGGTAAACAGACCCCGCCGCGCTACGACACCTCTTGTGTGTTTC 2393
QY 781 AspTyArgGlySerGlySerAspAlaSerLeuSerSerLeuThrSerSerAlaSer 800
DB 2394 GACTATGAGGCGAGCGGTCCGACCGCGCTCCCTGAGCTCCCTCAGCTCCCTCCGCTCC 2453
QY 801 AspGlnAspGlnAspTyArgPheLeuAsnGluTrpGlySerArgPheLysLysLeuAla 820
DB 2454 GACCAAGACCAAGATTACGATTATCTGAACGATGGGCGAGCCGCTTCCAGAGCTGGCA 2513
QY 821 AspMetTyArgGlyGlyGluAspAsp 829
DB 2514 GACATGTACGTGGCGGGGAGGACGAC 2540

RESULT 6

US-10-172-118-669
; Sequence 669, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; PRIOR APPLICATION NUMBER: 2002-06-14
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699


```
Db 1194 ACCACCCCTGAGAGAACCCAGGGCATCTGACACCCAGGAGGGTTTGGATTTTGGAGGCC 1253
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAACCAGCACACCCCTGCTAGCTTGAAGTACCAACAGGAGCCCTTTTGTGCTGAAGCTC 1313
Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCACCATAGTGGTCCACGTGGAGGATGTGAATGAGGACCTCTGTGTTT 1373
Qy 441 ValProProSerLysValValGluValGlnGluGlyLeuProThrGlyGluProValCys 460
Db 1374 GTCCACCCCTCCAAAGCTGTGAGGTCCAGAGAGGCAATCCCACTGGGGAGCCCTGTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTGCAGAGACCCCTGACAAAGAGAGAAATCAAAAGATCAGCTACCGCATCTCTGAGA 1493
Qy 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCACAGAGGTGGCTAGCCATCGACCCAGACAGTGGGAGGTGACAGCTGTGGGACCC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGAGGATGAGCAGTTTGTGAGGAACAACATCTATGAAGTCATGGTCTTGGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeuAsp 540
Db 1614 ATGACAAATGAAGCCCTCCACACCTGCGACGCGGAACCCCTCTGCTTAACACTGATTCAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGGCCAGTCCCTGAGCCCGTCAGATCACCATCTGCAACCAAGCCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGGCCACGTGTGTAACATCACGACAGACCTGTCTCCCAACACCTCCCTCTTCCAG 1793
Qy 581 AlaGlnLeuThrAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 1794 GCCAGCTCAGATGACTAGACATCTATGACGCGCAGGTCAACGAGGAAGGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGTCTTGTCCCTGGAAGAGTTCCTCAAGCAGGATACATATGACGTGCACCTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTGACCATGGCAACAAGAGCAGCTGACGTGATCAGGGCCACTGTGTCCGACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu 660
Db 1974 CATGGCCATGTGCAAACTGCCCTGGACCTCGAAGAGAGGTTCATCCTCCCTGTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2034 GGGGCTGTCTGTGCTCTGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2093
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAGATCAAGAGCCCTCTACTCCCAAGAGATGACACCCGTGACACAGTCTTCTAC 2153
Qy 701 TyrGlyGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCAAGAGGGGGTGGCAAGAGGACAGGACTATGACATCACCAGCTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db 2214 GGTCTGAGGCCAGGCGGAGGTGTCTCCGCAATGACGTGGCAACCAACCATCTCCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleLeuGlu 760
```

```
Db 2274 ACACCATGTACCGTCTAGCGCCAGCAACCCAGATGAAATCGCAACTTTTATAATTGAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGGGGCTTAACACAGACCCCAAGCCCGCCCTACGACACCTCTTGGTGTTC 2393
Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGCGAGCGGCTCCGACGCGCGTCCCTGAGCTCCCTCACCTCTCCGCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGTTATCTGAACGATGGGGCAGCGCGCTTCAAGAAGCTGGCA 2513
Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 2514 GACATGTACCGTGGCGGGAGGAGCGAC 2540

RESULT 8
US-10-158-123-2
; Sequence 2, Application US/10158123
; Publication No. US20030194406A1
; GENERAL INFORMATION:
; APPLICANT: Reinhard, Christoph
; APPLICANT: Klinger, Julie
; APPLICANT: Jefferson, Ann
; APPLICANT: Escobedo, Jaime
; APPLICANT: Randazzo, Filippo
; APPLICANT: Winter, Jill
; APPLICANT: Goodson, Robert
; APPLICANT: Qi, Weimin
; TITLE OF INVENTION: P-Cadherin as a Target for Anti-Cancer
; TITLE OF INVENTION: Therapy
; FILE REFERENCE: 35784/258994
; CURRENT APPLICATION NUMBER: US/10/158,123
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: 60/294,225
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 3171
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-158-123-2

Alignment Scores:
Pred. No.: 0 Length: 3171
Score: 4369.00 Matches: 829
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-09-916-849A-1 (1-829) x US-10-158-123-2 (1-3171)
Qy 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln 20
Db 54 ATGGGGCTCCCTCGTGGACCTCTCGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113
Qy 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
Db 114 TCGCGGGCTCCGAGCCCTCGCGCGTCTCTCAAGGAGGCTGAAGTACCTTGGAGCGC 173
Qy 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db 174 GGAGCGCGGAGCAGGACCCCGCCAGCGCTGGGGAAAGATTTCATGGGCTGCCCTGGG 233
Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyLysThr 80
Db 234 CAAGAGCCAGCTCTGTGTAGCACTGATAATGATCACTTCACTGTGCGGAATGGCGAGACA 293
Qy 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
```

294 GTCCAGGAAAGAGTCACTGAAGGAAGAGTCACTTGAAGATCTTCCCATCCAAACGT 353
 101 IleLeuArgArgHisLysArgAspTyrValValAlaProIleSerValProGluAsnGly 120
 354 ATCTTACGAAGACAAAGAGATGGTGGTGGTCTCAATATCTGCTCCCTGAATGGC 413
 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
 414 AAGGGTCCCTTCCCCAGAGACTGAATCAGCTCAAGTCAATTAATAAGATAGAGACCAAG 473
 141 IlePheTyrSerIleThrGlyProGlyValAlaAspSerProGluGlyValPheAlaVal 160
 474 ATTTTCTACAGCATCAGGGGCCCGGGGCGAGACAGCCCCCTGAGGGTGTCTTCGGCTGTA 533
 161 GlnLysGluThrGlyTyrLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
 534 GAGAGGAGACAGGCTGGTGTGTGTGAATAAGCCAGCTGACCGGGAGGAGATTGCCAAG 593
 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
 594 TATGAGCTCTTTGGCCAGCTGTGTACAGAAATGGTGCCTCAGTGGAGGACCCCATGAAC 653
 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
 654 ATCTCCATCATCGTACCGACCAAGATGACCAACAGCCCAAGTTTACCCAGGACACCTTC 713
 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
 714 CGAGGAGTGTCTTAGAGGAGTCTTACCAGGTACTTCTGTGTGTCAGGTGACAGCCACA 773
 241 AspGluAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
 774 GATGAGGATGATGCACTTACCTACCTACATGCGGTGGTGTCTTACTTCATCCATAGCCAA 833
 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
 834 GAACCAAGGACCCACACGACCTCATGTTCACATTCACCGAGCAGACGACCATCAGC 893
 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrIleThrIleGlnAla 300
 894 GTCATCTCCAGTGGCTCGACCGGAAAGTCTCCTGAGTACACACTGACATCCAGGCC 953
 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValAlaValGluIleLeuAsp 320
 954 ACAGACATGGATGGGACGGCTCCACCACCGCAGTGGCGAGTAGTGGAGATCCTTGAT 1013
 321 AlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
 1014 GCCAATGCAATGTCTCCCATGCTTTGACCCCGAGAGTACGAGGCCCATGTGCTCGAGAAT 1073
 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
 1074 GCAGTGGGCCCATGAGTGCAGAGGCTGACGCTGACTGATCTGGAGGCCCCCACTCACCA 1133
 361 AlaThrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
 1134 GCGTGGCGTGCACCTATCTATATGCGCGGTGACGAGGGGCCCATTTTACCATCACC 1193
 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
 1194 ACCCACTTGAGAGCAACAGGCGCATCTGTCAACACGAGGAGGTTTGGATTTGAGGCC 1253
 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
 1254 AAAAACCAGCACACCTGTAGTGAAGTGAAGTGAACCAAGGCGCCCTTTTGTGCTGAAGCTC 1313
 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
 1314 CCAACCTCCAGGACCATAGTGTGTCACGTGGAGGATGTGATGAGGACCTGTGTTT 1373
 441 ValProProSerLysValValGluValGlnGluIleProThrGlyGluProValCys 460

1374 GTCCACACCTCCAAAGTGGTTGAGGTCCAGGAGGGCATCCCACTGGGGAGCCCTGTGTGT 1433
 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
 1434 GTCTACACTGCAGAGACCCCTGACAAGGAGAAATCAAAAGATCAGCTACCGCATCCTGAGA 1493
 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
 1494 GACCCAGCAGGGTGGCTAGCCATGGACCAACAGTGGGAGGTGACAGTGTGGGCACC 1553
 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
 1554 CTCGACCGTGGATGAGCAGTTTGTGAGGAACAACATCTATGAAGTCATGCTTGGCC 1613
 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuLeuLeuLeuLeu 540
 1614 ATGGCAATGGAAGCCCTCCACCATGGCAGGAAACCTTCTGCTAACACTGATTGAT 1673
 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
 1674 GTCAACGACCATGGCCCACTCCCTGAGCCCCGTGAGTACCATCTGCAACCAAGCCCT 1733
 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
 1734 GTCGCGCACGTGCTGAACATCAGGACAAAGACCTGTCTCCCCACACCTCCCTTTCAG 1793
 581 AlaGlnLeuThrAspAspSerAspIleTyrTyrThrAlaGluValAsnGluGluGlyAsp 600
 1794 GCCCAGCTCAGATGACTCAGACATCTACTGGACGGCAGAGGTCAACGAGGAAGTGAC 1853
 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
 1854 ACAGTGGTCTTCTCCCTGAAGAAGTTCTGAGCAGGATACATATGACGTGCACCTTCT 1913
 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
 1914 CTGTCTGACCATGGCAACAAGAGAGCTGACGTGATCAGGGCCACTGTGTGGACTGC 1973
 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu 660
 1974 CATGGCCATGTGCAACCTGCCTCGAACCTGGAAAGGAGGTTTCATCTCCCTGTGTGTG 2033
 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuLeuValArgLysLys 680
 2034 GGGGTGTCTCGCTGTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2093
 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
 2094 CGGAAGATCAAGAGGCCCTCTCTACTCCAGAAAGATGACACCCGTGACACAGTCTCTAC 2153
 701 TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnHisArg 720
 2154 TATGGCAAGAGGGGGTGGCAAGAGGACCAAGGACTATGACATCACCAGCTCCACCGA 2213
 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
 2214 GGTCTGAGGCCAGGCCGCGAGGTGTTCTCCGCAATGACGTGGCACCACCATCATCCCG 2273
 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
 2274 ACACCCATGTACCTCTAGGCCAGCCACCAACCCAGATGAATCGGCAACTTATAATTGAG 2333
 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
 2334 AACCTGAAGCGGCTTAACACAGCCCAAGCCCGCCCTAGCACACCTCTTGTGTGTT 2393
 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
 2394 GACTATGAGGCGAGCGCTCCGACCGCGTCCCTGAGTCCCTCCTCCTCCTCCTCCTCCTCC 2453
 801 AspGlnAspGluAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
 2454 GACCAAGACAGATATACGATTATCTGACAGAGTGGGCGAGCGCTTCAAGAGAGTGGCA 2513

Db	174	GGAGCGCGGAGCAGAGCCGGCCAGCGCTGGGGAAAGTAATTATGGCTGCCTCGG	233
Qy	61	GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr	80
Db	234	CAAGAGCCAGCTCTGTTTAGCAGCTGATATGATGATCTTCACTGTGGGAATGCGGACA	293
Qy	81	ValGlnGluArgArgSerLeuLeuGlyGluArgAsnProLeuLysIlePheProSerLysArg	100
Db	294	GTCCAGGAAAGAGTCACTGAAGAAAGGAATCCATTGAAGATCTTCCATCCAAACGT	353
Qy	101	IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly	120
Db	354	ATCTTACCAAGACACAAGAGAGATTGGTGGTGTCTCAATATCTGTCCCTGAAATGCG	413
Qy	121	LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys	140
Db	414	AAGGGTCCCTTCCCCCAGAGACTGAATCAGCTCAAGTCTAATAAGATAGAGACCAAG	473
Qy	141	IlePheTyrrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal	160
Db	474	ATTCTTACAGCATCACGGGCCCGGGCAGACAGCCCGCTGAGGGTGTCTTCCTGTA	533
Qy	161	GluLysGluThrGlyTyrrPheLeuLeuAsnLysProLeuAspArgGluGluLeuAlaLys	180
Db	534	GAGAAGGAGACAGCGCTGGTCTTGTGAATAAGCCACTGGACCGGGAGAGATTGCCAAG	593
Qy	181	TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn	200
Db	594	TATCAGCTCTTTGGCCACCGCTGTCTCAGAGAATGGTGCCTCAGTGGAGACCCCATGAAC	653
Qy	201	IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe	220
Db	654	ATCTCCATCATCGTGACCGACCAAGATGACCACAAAGCCCAAGTTTACCAGGACACCTTC	713
Qy	221	ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr	240
Db	714	CGAGGGAGTGCTTAGAGGGAGTCCCTACAGGTACTTCTGTGATGCAGGTGACAGCCACA	773
Qy	241	AspGluAspAspAlaIleTyrrThrTyrrAsnGlyValValAlaTyrrSerIleHisSerGln	260
Db	774	GATGAGGATGATGCCATCTACACCTACATGGGGTGGTGTGCTTACTCATCATAGCCAA	833
Qy	261	GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer	280
Db	834	GAACCAAGACCCACACAGCACTCATGTTCCAAATTCACGGAGCAGGACCATCAGC	893
Qy	281	ValIleSerSerGlyLeuAspArgGluLysValProGluTyrrThrLeuThrIleGluAla	300
Db	894	GTCATCTCCAGTGGCCTGGACCGGAAAAGTCCCTGAGTACACACTGACCATCCAGGCC	953
Qy	301	ThrAspMetAspGlyAspGlySerThrThrThraValaValaValGluIleLeuAsp	320
Db	954	ACAGACATGATGGGACGGCTCCACCACACCGCAGTGGCAGTAGTGGAGATCCTTGAT	1013
Qy	321	AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrrGluAlaHisValProGluAsn	340
Db	1014	GCCAAATGACATGCTCCCATGTTTGACCCCCAGAGATGACGGGCCCATGTGCCTAGAAAT	1073
Qy	341	AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro	360
Db	1074	GCASTGGGCCATGAGGTGACAGGCTGACGGTCACTGATCTGAGCGCCCACTCACCA	1133
Qy	361	AlaTrpArgAlaThrTyrrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr	380
Db	1134	GCGTGGCGTGCCACCTTACCTTATCATGGGGGTGACGACGGGGACCCATTTTACATCACC	1193
Qy	381	ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla	400
Db	1194	ACCCACCTGTAGAGCAACACGGGCATCCTGTACAAACAGGAAGGGTGTGGATTTTGAGGCC	1253
Qy	401	LysAsnGlnHisThrLeuTyrrValGluValThrAsnGluAlaProPheValLeuLysLeu	420

Db 1254 AAAAAACAGCAGACCCCTGACGTGAAGTGAACCAAGAGCCCTTTTGTGCTGAGCTC 1313
 Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
 Db 1314 CCAACCTCCACAGCCACCATAGTGTCCACGTGGAGGATGTAATGAGGCACCTGTGTT 1373
 Qy 441 ValProProSerIysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
 Db 1374 GTCCCAACCTCCAAAGTCGTGAGTCCAGGAGGACATCCCACTGGGAGCCCTGTGTGT 1433
 Qy 461 ValThrThrAlaGluAspProAspIysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
 Db 1434 GTCTACATGCAGAACCCCTGCACAGGAGATCAAAAGATCAGCTACCGCATCTCGAGA 1493
 Qy 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
 Db 1494 GACCCAGCAGGGTGGTGTAGCCATGAGCACCAGACAGTGGGCAAGGTACAGCTGTGGGCACC 1553
 Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
 Db 1554 CTCGACCGTGAGGATGAGCAGTTTGTGAGGAACAACATCTATGAAGTCATGCTTGCC 1613
 Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
 Db 1614 ATGGACAATGGAGGCCCTCCCACTGCGACCGGAACCTTCTGCTAACCACTGATTGAT 1673
 Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
 Db 1674 GTCAACGACCATGGCCCTCCCTGAGCCCTGCATCACCATCTGCACCAACCAAGCCCT 1733
 Qy 561 ValArgHisValLeuAsnIleThrAspIysAspLeuSerProHisThrSerProPheGln 580
 Db 1734 GTGGCGACATGCTGAACATCAAGCAAGGACCTGTCTCCCAACACCTCCCTCTCCAG 1793
 Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
 Db 1794 GCCCAGCTCACAGATGACTCAGACATCTACTGGACGGCAGAGGTCAACGAGGAAGTGAC 1853
 Qy 601 ThrValValLeuSerLeuIysIysPheLeuIysGlnAspThrTyrAspValHisLeuSer 620
 Db 1854 ACAGTGTCTTGTCCCTGAAGAGTTCCTGAGCAGGATACATATGATGACGTGCACCTTCT 1913
 Qy 621 LeuSerAspHisGlyAsnIysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
 Db 1914 CTGTCTCAGCATGGCAACAAGAGCAGTGCAGGTGATCAGGCGCATCTGTGCGACTGC 1973
 Qy 641 HisGlyHisValGluThrCysProGlyProThrIysGlyPheIleLeuProValLeu 660
 Db 1974 CATGGCCATGTGCAAACTCCCTGACCCCTGGAAGAGGATTTTCATCCTCCCTGCTGCTG 2033
 Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
 Db 2034 GGGGCTGTCTGCTCTGCTGTCTCTCTGCTGTGCTGTCTTTTGTGTTGAGAAAGAG 2093
 Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
 Db 2094 CGGAAGATCAAGAGCCCTCTCTCTCCAGAGATGACCCCGTGAACAGCTCTCTCTAC 2153
 Qy 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
 Db 2154 TATGGCGAAGAGGGGTGGCGAAGAGGACCATGATGATGATGATGATGATGATGATGAT 2213
 Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
 Db 2214 GGTCTGGAGCCAGCCCGAGGTGTTCTCCGCAATGAGCGGACCAACCAATCATCCCG 2273
 Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleLeuGlu 760
 Db 2274 ACACCCATGTACCGTCTTAGCCACCAACCCAGATGAATCGCACTTATATATGAG 2333
 Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
 Db 2334 AACCTGAAGGGCGGTAAACACAGACCCCAACCCCGCCCTACGACACCCCTCTTGTGTGTC 2393

Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
 Db 2394 GACTATGAGGCGAGGCTCCGACGCGGTCCCTGAGTCCCTCACCTCTCGGCTCC 2453
 Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheIysLeuAla 820
 Db 2454 GACCAAGACCAAGATACGTTATCTGAACAGTGGGCGACCGCTTCAAGAAGCTGGCA 2513
 Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
 Db 2514 GACATGTACGTGGCGGAGGACGAC 2540

RESULT 10
 US-10-295-027-799
 ; Sequence 799, Application US/10295027
 ; Publication No. US20030232350A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Afar, Daniel
 ; APPLICANT: Aziz, Natasha
 ; APPLICANT: Ginsberg, Wendy M.
 ; APPLICANT: Gish, Kurt C.
 ; APPLICANT: Glynn, Richard
 ; APPLICANT: Hevezi, Peter A.
 ; APPLICANT: Mack, David H.
 ; APPLICANT: Murray, Richard
 ; APPLICANT: Watson, Susan R.
 ; APPLICANT: Bos Biotechnology, Inc.
 ; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
 ; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
 ; FILE REFERENCE: 018501-012500US
 ; CURRENT APPLICATION NUMBER: US/10/295,027
 ; CURRENT FILING DATE: 2002-11-13
 ; PRIOR APPLICATION NUMBER: US 09/663,733
 ; PRIOR FILING DATE: 2000-09-15
 ; PRIOR APPLICATION NUMBER: US 60/350,666
 ; PRIOR FILING DATE: 2001-11-13
 ; PRIOR APPLICATION NUMBER: US 60/335,394
 ; PRIOR FILING DATE: 2001-11-15
 ; PRIOR APPLICATION NUMBER: US 60/332,464
 ; PRIOR FILING DATE: 2001-11-21
 ; PRIOR APPLICATION NUMBER: US 60/334,393
 ; PRIOR FILING DATE: 2001-11-29
 ; PRIOR APPLICATION NUMBER: US 60/340,376
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: US 60/347,211
 ; PRIOR FILING DATE: 2002-01-08
 ; PRIOR APPLICATION NUMBER: US 60/347,349
 ; PRIOR FILING DATE: 2002-01-10
 ; PRIOR APPLICATION NUMBER: US 60/355,250
 ; PRIOR FILING DATE: 2002-02-08
 ; PRIOR APPLICATION NUMBER: US 60/356,714
 ; PRIOR FILING DATE: 2002-02-13
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 1386
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 799
 ; LENGTH: 3171
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-295-027-799

Alignment Scores:
 Pred. No.: 0 Length: 3171
 Score: 4369.00 Matches: 829
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 16 Gaps: 0

US-09-916-849A-1 (1-829) x US-10-295-027-799 (1-3171)
 Qy 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuGlnValCysTrpLeuGln 20

[illegible]

1134	Db	CGCTGGCGTGCACATCTACCTTATCATGGCGGGTGACGACGGGGACCATTTTACCATCACCC	1193
381	Qy	ThrHisProGluSerAsnGlnGlyVlleIeuThrThrArgIysGlyLeuAspPheGluAla	400
1194	Db	ACCCACCCCTGAGACCAACCGAGGCATCCTGACACACAGGAGGGTCTTGGATTTTGGAGGCC	1253
401	Qy	LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu	420
1254	Db	AAAAACAGCACACCCCTGTAGTTGAAGTGACCAACGAGGCCCTTTTGTGCTGAAGCTC	1313
421	Qy	ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe	440
1314	Db	CCAACTCCACAGCCACCATAGTGGTCCACGTGGAGGATGTGAATGAGGCACCTGTGTTT	1373
441	Qy	ValProSerLysValValGluValGlnGlnGlyIleProThrGlyGluProValCys	460
1374	Db	GTCCACCCCTCCAAAGTCGTTGAGTCCAGAGGGGCATCCCCACTGGGAGACCTGTGTGT	1433
461	Qy	ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg	480
1434	Db	GTCTACACTCGAAGACCCCTGCAAGAGGAATCAAAAGATCAGTCACGCATCCTGAGA	1493
481	Qy	AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr	500
1494	Db	GACCCAGCAGGTGGCTAGCCATGGACCCACAGTGGCGAGTCAAGCTGTGGGCACC	1553
501	Qy	LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla	520
1554	Db	CTCGACCGTGAAGATGAGCAGTTTGTGAGGAACAACATCTATGAAGTCATGGTCTTGGCC	1613
521	Qy	MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrIleLeuAsp	540
1614	Db	ATGGACATGGAGGCCCTCCACACTGGCACGGGAACCCCTCTGCTTAACACTGATTGAT	1673
541	Qy	ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro	560
1674	Db	GTCAACGACCATGGCCCGAGTCCCTCGAGCCCGTCAGATCACCATCTGCACCAAGGCCCT	1733
561	Qy	ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln	580
1734	Db	GTGGCCACGTGCTGAACATCAGCGACAGGACCTGTCTCCCCACACCTCCCTTCCACG	1793
581	Qy	AlaGlnLeuThrAspSerAspIleTyrTyrThrAlaGluValAsnGluGluValAsp	600
1794	Db	GCCAGCTCACAGATGACTCAGACATCTACTGGACGGCAGAGGTCAACGAGGAGGTGAC	1853
601	Qy	ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer	620
1854	Db	ACAGTGGTCTTGTCCTGAAGAAGTTCTTGAAGCAGGATACATATGACGTGCACCTTTCT	1913
621	Qy	LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys	640
1914	Db	CTGTCTGACCATGGCAACAAGAGCAGCTGACGGTGATCAGGGCCACTGTGTGCGACTGC	1973
641	Qy	HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu	660
1974	Db	CATGGCCATGTGAAACCTGCCCTGGACCCCTGAAAGGAGGTTTCATCTCCCTGTGCTG	2033
661	Qy	GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys	680
2034	Db	GGGGCTGTCTGGCTCTGTGTCTCTCTCTGTGGTGCTCTCTTTTGTGTGTGAAGAAG	2093
681	Qy	ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr	700
2094	Db	CGAAGATCAAGAGAGCCCTCTACTCTCCAGAAGATGACACCGGTGACACGCTCTCTAC	2153
701	Qy	TyrGlyGluGlnGlyGlyGluGluAspGlnAspTyrAspIleThrClnIleuHisArg	720
2154	Db	TATGCCAAGAGGGGGTGCCGAAGAGGACCAAGGACTATGACATCACCAGCTCCACCGA	2213
721	Qy	GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro	740
2214	Db	GGTCTGGAGCCAGGCCGAGGTGGTTCTTCCCAATGACGTGGCACCAACCATATCCCG	2273

Db 1014 GCCAATGCAATGCTCCCATGTTTGACCCCCAGAGTACGAGGCCCATGTGCTGAGCAAT 1073
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1074 GCAGTGGCCCATGAGGTGCAGAGGTGACGGTCACTGATCTGGAGCGCCCACTACCA 1133
Qy 361 AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspGlyAspHisPheThrIleThr 380
Db 1134 GCGTGGCGTCCACCTACCTATATCATGGCGGTGACGACGGGACCACTTTTACCATCACC 1193
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1194 ACCACCCCTGAGACCAACAGGGCATCTGCAACACAGGAGGGTTTGGATTTTGGAGCC 1253
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAAACAGACACACCTGTGAGTTGAAGTGACCAACAGGCCCCCTTTTGTGTGAAGCTC 1313
Qy 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCCACCATAGTGGTCCACGTGAGGATGTGAATGAGGCACCTGTGTTT 1373
Qy 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1374 GTCCACCCCTCCAAAGTGTGTGAGGTCCAGAGGGCATCCCACTGGGGAGCCTGTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTGCAGAGAACCCCTGACAGAGAGAAATCAAAAGATCAGTACCGCATCCTGAGA 1493
Qy 481 AspProIaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCACAGGGTGGCTAGCCATGACCCAGACAGAGTGGGAGGTGACAGTGTGGGCACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTGACCGTGGAGTGAAGTGTGTGAGGAACAACATCTATGAGTCATGCTTGGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db 1614 ATGGACAATGAGACCCCTCCACCACCTGCGACGGACGGAAACCCCTTCTGCTAACTGATTAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGCCCGAGTCCCTGAGCCCCGTGATCACCATCTGCAACCAAGCCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGGCCACGTGCTGAACATCAAGGACAGACGACCTGTCTCCACACCTCCCTTTCAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 1794 GCCAGCTCAGATGATGACTCAGACATCTACTGGACGCGAGGTCAACGAGGAAGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGTCTGTCCCTGAGAGAGTTCCTGACGACGATACATATGACGTGCACCTTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTGACCATGGCAACAAGAGCAGCTGACGGTGTATCAGGGCCACTGTGTGCGACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTyrPheLeuPheIleLeuProValLeu 660
Db 1974 CATGGCCATGTGNAACCTGCCCTGGACCTCGAAGAGAGGTTCATCTCCCTGCTGTG 2033
Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2034 GGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2093
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAAGGAGGCCCTCTCTACTCCCAAGAGATGACACCCGTGACAAACGCTTCTAC 2153

Qy 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCGAAGAGGGGGTGGCGAAGAGGACCAAGGACTATGACATCACCCAGCTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db 2214 GGTCTGGAGGCCAGCGCGAGGTGTCTCCGCANTGAGCTGGCACCAACCATCATCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db 2274 ACACCCATGATCGCTAGGCGAGCAACCCAGATGAATCGCAACTTTTATAATTGAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGCGCGCTTAACAGACCCACGCCCCCTACACACCTCTTGTGTTC 2393
Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGCGAGCGGCTCCGACGCGGTCCCTGAGCTCCCTCACCTCTCCGCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGGGAGCCGCTTCAAGAAGCTGGCA 2513
Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 2514 GACATGTACGCTGGCGGGGAGGACGAC 2540

RESULT 12

US-10-2995-027-895
; Sequence 895, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynnne, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 895

QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgIysGlyLeuAspPheGluAla 400
DB 1194 ACCCACCTGAGAGCAACAGGCGATCTGCAACACAGAGAGGTTTGGATTGTAGGCC 1253
QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
DB 1254 AAAAACACGACACCTGTGAGTGTGAGTGCACCAACGAGGCCCTTTGTGTGAGCTC 1313
QY 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
DB 1314 CCAACCTCCACAGCACCATAGTGTCCACGTGAGGATGTAATGAGGACCTGTGTTT 1373
QY 441 ValProProSerLysValValGluValGlnGlyIleProThrGlyGluProValCys 460
DB 1374 GTCCACACCTCCAAAGTCGTTGAGTCCAGAGGCGCATCCCACTGGGAGCCTGTGTGT 1433
QY 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
DB 1434 GTCTACACTGCAGAAGACCTTGACAGAGGAGTAATCAAAAGATCAGCTACCGATCCTGAGA 1493
QY 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
DB 1494 GACCCAGCAGGCTGCTAGCATGACCCACAGACAGTGGCAGGTCAAGCTGTGGGCACC 1553
QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
DB 1554 CTCGACCTGAGGTAGCAGCTTTGTGAGGACACATCTATGAATCATGGTCTTGCC 1613
QY 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
DB 1614 ATGGACAATGGAAGCCCTCCACCACTGGCACCGGAAACCTTCTGCTAAACATGATGAT 1673
QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
DB 1674 GTCAACGACCATGGCCCGAGTCCCTGAGCCCGCTGATCAACATCTGCAACCAAGCCCT 1733
QY 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
DB 1734 GTGGCCACGCTGTCAACATCACGACCAAGACCTGTCTCCACACACTCCCTTTCACG 1793
QY 581 AlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGlyAsp 600
DB 1794 GCCCAGCTCACAGATGACTAGACATCTACTGACGCGGACAGGTCACAGGAGAGTGAC 1853
QY 601 ThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
DB 1854 ACAGTGGTCTTGCTCCCTGAAGAGTTCCTGAAGCAGGATACATATGACGTGACCTTCT 1913
QY 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
DB 1914 CTGTCTGACCATGGCAACAAAGACAGCTGACGCTGATCAGGGCCACTGTGTGCGACTGC 1973
QY 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu 660
DB 1974 CATGCCCATGTGCAAACTCGCTGGACCTTGAAAGAGGTTTCACTCCCTGCTGCTG 2033
QY 661 GlyAlaValLeuAlaLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
DB 2034 GGGGCTGTCTGCTGCT 2093
QY 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
DB 2094 CGAAGATACAGAGCCCTCTACTCTCCAGAGATGACACCCCGTACACAGCTCTCTCTAC 2153
QY 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
DB 2154 TATGCCGAAGAGGGGGTGGGAGAGGACAGGACCATGACATCAACCCAGCTCCACCGA 2213
QY 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
DB 2214 GGTCTGGAGGCGAGCCGAGTGTCTCTCGCAATGACGTGGCACCACCATCATCTCCG 2273
QY 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760

DB 2274 ACACCATGTACCTGCTTAGCCACCCACCATGAAATCGCACTTATATTTAG 2333
QY 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProTyrAspThrLeuLeuValPhe 780
DB 2334 AACCTGAAGGCGCTAACACACAGACCCACAGCCCGCCTACGACACCTCTTGTGTTC 2393
QY 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerAlaSer 800
DB 2394 GACTATGAGGCGAGCGGCTCCGACCCCGCTCCCTGAGCTCCCTCACCCTCTCCGCTCC 2453
QY 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGlnIleTyrGlySerArgPheLysLeuAla 820
DB 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCCGCTTCAAGAAGCTGGCA 2513
QY 821 AspMetTyrGlyGlyGlyGluAspAsp 829
DB 2514 GACATGTACGTGGCGGGGAGGACGAC 2540
RESULT 14
US-10-775-920-188
; Sequence 188, Application US/10775920
; Publication No. US2004017544A1
; GENERAL INFORMATION:
; APPLICANT: Mergen Ltd
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES AND CORRESPONDING ENCODED POLYPEPTIDES
; TITLE OF INVENTION: OF PARTICULAR SECRETED AND MEMBRANE-BOUND PROTEINS OVEREXPRESSED
; TITLE OF INVENTION: IN CERTAIN CANCERS
; FILE REFERENCE: Mergen - 0010B
; CURRENT APPLICATION NUMBER: US/10/775,920
; PRIOR FILING DATE: 2004-02-10
; PRIOR APPLICATION NUMBER: US 60/447,900
; NUMBER OF SEQ ID NOS: 385
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 188
; LENGTH: 3171
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-920-188
Alignment Scores:
Pred. No.: 0 Length: 3171
Score: 4369.00 Matches: 829
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0
US-09-916-849A-1 (1-829) x US-10-775-920-188 (1-3171)
QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuLeuValCysTrpLeuGln 20
DB 54 ATGGGCTCCCTCGTGAGACCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113
QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
DB 114 TCGCGGCTCTCGAGCCCTGCGCGGCTCTTCAGGAGGCTGAAGTACCTTGGAGCG 173
QY 41 GlyGlyAlaGluGlnProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
DB 174 GGAGGCGCGAGCAGGAGCCCGCCAGCGCTGGGAAAAGTATTCATGGCTGGCTGG 233
QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
DB 234 CAAGACCCAGCTCTGTTTATGACTGATATGATGATCTCTCTCTCTCTCTCTCTCTCT 293
QY 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
DB 294 GTCCAGAAAGAGGTCACTGAAGAAAGGAATCCATTGAAGATCTTCCCATCCAAAGCT 353
QY 101 IleLeuArgArgHisLysArgAspTrpValAlaProIleSerValProGluAsnGly 120

Db 354 ATCTACGACACACAGAGATTGGGTGGTCTCCAAATATCTGTCCCTGAAAATGGC 413
Qy 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
Db 414 AAGGGTCCTTCCCCGACAGACTGAATCAGCTCAAGTCTAATAAAGATAGACACCAAG 473
Qy 141 IlePheTyrSerIleThrGlyProGluValAspSerProGluGluValPheAlaVal 160
Db 474 ATTTTCTACACATCACGGGCGGGGACAGACCCCTGAGGGTGCTTCGCTGTA 533
Qy 161 GluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
Db 534 GAGAAGGAGACAGGCTGGTGTGTGAATAAGCCACTGGACCGGAGGAGATTGCCAAG 593
Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 594 TATGAGTCTTTGGCCACGCTGTGTACAGAAATGGTGCCTCAGTGGAGACCCCATGAC 653
Qy 201 IleserIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 654 ATCTCCATCATCGTGACCGACACAGATGACCACAGCCCAAGTTTACCCAGGACACCTTC 713
Qy 221 ArgGlySerValLeuGluGluValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 714 CGAGGGAGTGTCTAGAGGGAGTCTTACCAGGTACTTCTGTGATGCAGGTGACAGCCACA 773
Qy 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
Db 774 GATGAGGATGATGCATCTACCTCAATGGGGTGGTGTGTTACTCAATCCATAGCCAA 833
Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 834 GAACCAAGGACCCACACGACCTCATGTTTCAATTCACCGGAGCACAGGCACCATCAGC 893
Qy 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrIleuThrIleGlnAla 300
Db 894 GTCACTCCAGTGCGCTGGACCGGAAAAGTCCCTCAGTACACACTGACCATCCAGGCC 953
Qy 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValAlaValGluIleLeuAsp 320
Db 954 ACAGACATGATGGGACGGCTCCACCCACCGCAGTCAGTGTAGTGAGATCCTTGAT 1013
Qy 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1014 GCCAATGACATGCTCCCATGTTTGACCCCGAGAGTACGAGGCCCATGTGCGCTGAGAA 1073
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1074 GCAGTGGGCCATGAGGTGCAGAGCTGACGGTCACTGATCTGAGACGCCCCCACTACCA 1133
Qy 361 AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db 1134 CGGTGGCGTGCCACTTATCATGGCGGTGACGACGGGACCACTTTTACCATCACC 1193
Qy 381 ThrHisProGluSerAsnGlnGlyIleuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1194 ACCACCCCTGAGACACACAGGGATCTCTGACACACCGAGGAGGGTTTGGATTGTAGGCC 1253
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAAACGAGCACACCTGTGAGTTGAAGTACCAACGAGGCCCTTTTGTGCTGAGGCTC 1313
Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCCACCATAGTGGTCCAGTGGAGGATGTGAATGAGGACCTGTGTTT 1373
Qy 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1374 GTCCCAACCTCCAAAGTCGTGTAGGTCCAGAGGGGATCCCTCCTGGGGAGCCTGTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTGCAGAAAGCCCTGTACAGAGAGAAATCAAAAGATCAGCTACCGCATCTCTGAGA 1493

Qy 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCAGCAGGGTGGGTAGCCATGGACCCAGACAGTGGGCGAGTCCACAGCTGTGGCACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGGAGATGAGCAGTTTGTGAGAAACACATCATTAAGATCATGTCTTGGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeuAsp 540
Db 1614 ATGGCAATGGAAGCCCTCCACCACTGGACCGGAAACCTTCTGCTAAACATGATTGAT 1673
Qy 541 ValAsnAsnHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGGCCCGTCCCTGAGCCCTCAGATCACCATCTGCAACCAAGCCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGGCCACGTGCTGAACATCACGGACAGGACCTGTCTCCACACACCTCCCTTCCAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 1794 GCCCAGCTCACAGATGACTCAGACATCTACTCGACGGCAGAGGTCAACGAGAAAGGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGGTCTTCCCTGAAAGTTCCTGAAGCAGGATACATATGACGTGACCTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGCTCGACCATGGCAACAAAGAGAGCTGACGGTGATCAGGGCCACTGTGTGCGACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleuProValLeu 660
Db 1974 CATGGCCATGTGAAACCTGCCCTGACCTCGAAGAGGAGTTTCATCTCCTCTGTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2034 GGGGTGTCTCTGCTGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2093
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAAGAGGCCCT 2153
Qy 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCCAGAGAGGGGGTGGCGGAGAGGACGAGGACTATGACATCACCCAGCTCCACCA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
Db 2214 GGTCTGGAGGCCAGGCGGCGGAGGTGTCTCTCGCAATGACGTGGCACCACCACTATCCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleLeuGlu 760
Db 2274 ACACCATGTACCGTCTTAGGCCAGCAACCCAGATGAAATCGGCACTTTTATAATTGAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGGGCGCTAAACACAGACCCACAGCCCGCCCTACGACACCTCTTGTGTTC 2393
Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGGGAGCGGCTCCGACCGCGCTCCCTGAGCTCCCTCAGCTCTCCGCCCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCGCTTCAAGAGCTGGCA 2513
Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 2514 GACATGTACGTGGCGGGGAGGACGAC 2540

Db	2296	GTGGCCACGTCGTGAACATCACGGAAGACCTGTCTCCACACCTCCCTTTCCAG	2355
Qy	581	AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGluGlyAsp	600
Db	2356	GCCAGCTCACAGATGACTCAGACATCTACTGGACGCGCAGAGGTCAACGAGGAAGTGAC	2415
Qy	601	ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer	620
Db	2416	ACAGTGGTCTTGTCCCTGAAGAAGTCTCTGAAGCAGGATACATATGACGTGCACCTTTCT	2475
Qy	621	LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys	640
Db	2476	CTGTCTGACCATGGCAACAAGAGCAGCTGACGGTGATCAGGGCCACTGTGTGCACTGC	2535
Qy	641	HisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValLeu	660
Db	2536	CATGGCCATGTCGAAACCTGCTCCCTGGACCCCTGGAAAGGAGGTTCATCCTCCTGTCTGT	2595
Qy	661	GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuValArgLysLys	680
Db	2596	GGGGCTGTCTGGCTCTGCTGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2655
Qy	681	ArgLysLysGluProLeuLeuLeuProGluLeuAspThrArgAspAsnValPheTyr	700
Db	2656	CGGAAGATCAAGGAGCCCTCTACTCCAGAGATGACACCCGTGACAAAGTCTTCTAC	2715
Qy	701	TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg	720
Db	2716	TATGGCAAGAGGGGGGGGGAGAGAGCAGGACTATGACATCACCAGCTCCACCGA	2775
Qy	721	GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro	740
Db	2776	GGTCTGGAGGCCAGCGCGAGGTGGTCTCTCGCATGACGTGGCACCACCATCATCCCG	2835
Qy	741	ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu	760
Db	2836	ACACCCATGTACCGTCTTAGCCAGCCCAACCCAGATGAAATCGGCAACTTTATAATTGAG	2895
Qy	761	AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuValPhe	780
Db	2896	AACCTGAAGCGGGTTACACAGACCCCAAGCCCGCCCTACGACACCCCTCTGGTGTTC	2955
Qy	781	AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer	800
Db	2956	GACTATGAGGCGAGCGGCTCCGACGCGCGTCCCTGAGTCCCTCACCTCCTCGCCTCC	3015
Qy	801	AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLysLeuAla	820
Db	3016	GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGGCAGCCGCTTCAAGAAGCTGGCA	3075
Qy	821	AspMetTyrGlyGlyGlyGluAspAsp	829
Db	3076	GACATGTACGTGGGGGGGAGGACGAC	3102

Search completed: September 23, 2004, 03:46:12
Job time : 1023 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:03:42 ; Search time 21 Seconds
(without alignments)
3797.275 Million cell updates/sec

Title: US-09-916-849A-1
Perfect score: 4369
Sequence: 1 MGLPRGPLASLLLLQVCWLQ.....NEWGRFKKLADMYGGEDD 829

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.78.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4369	100.0	829	1 IJHUCP	cadherin 3 precurs
2	3555.5	81.4	822	1 IJMSCF	P-cadherin precurs
3	2573.5	58.9	732	1 IJHCHB	B-cadherin precurs
4	2420	55.4	882	1 IJHUCE	cadherin 1 precurs
5	2388	54.7	884	1 IJMSCE	E-cadherin precurs
6	2379	54.5	884	2 S34438	uvomorulin - mouse
7	2375	54.4	895	1 IJXLCP	EP-cadherin precurs
8	2371	54.3	905	2 S43064	cadherin - African
9	2323	53.2	491	1 IJBOCP	P-cadherin - bovin
10	2274	52.0	887	1 IJCHCL	cadherin precurs
11	2192.5	50.2	871	2 S47518	cadherin - African
12	1707	39.1	905	1 IJHUCN	cadherin 2 precurs
13	1704.5	39.0	912	1 IJHUCN	N-cadherin precurs
14	1697.5	38.9	906	1 IJMSCH	N-cadherin precurs
15	1694	38.8	913	1 IJHCHC	R-cadherin precurs
16	1691	38.7	877	1 IJBOCN	N-cadherin precurs
17	1655	37.9	913	1 A47543	R-cadherin precurs
18	1621.5	37.1	916	2 C38992	cadherin 4 precurs
19	1621	37.1	783	2 I50116	N-cadherin precurs
20	1616.5	37.0	905	1 IJXLG2	N-cadherin 2 precu
21	1609	36.8	905	1 IJXLG1	N-cadherin 1 precu
22	1321.5	30.2	814	2 G02878	cadherin-15 precu
23	1302.5	29.8	730	1 IJMSCH	M-cadherin - mouse
24	1021.5	23.4	713	2 B38992	cadherin 13 precu
25	1007.5	23.1	785	2 I50180	cadherin-7 - chick
26	990	22.7	790	2 I50178	cadherin-6B - chic
27	982	22.5	790	2 I37016	cadherin-6 - huma
28	976.5	22.4	790	2 G02678	cadherin-14 - huma
29	976	22.3	712	1 IJMSCT	T-cadherin precurs

RESULT 1

IJHUCP

cadherin 3 precursor - human

N;Alternate names: P-cadherin; placental cadherin

C;Species: Homo sapiens (man)

C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999

R;Accession: A33659

R;Shimoyama, Y.; Yoshida, T.; Terada, M.; Shimosato, Y.; Abe, O.; Hirahashi, S.

J. Cell Biol. 109, 1787-1794, 1989

A;Title: Molecular cloning of a human Ca(2+)-dependent cell-cell adhesion molecule homol

A;Reference number: A33659; MUID:90009051; PMID:2793940

A;Accession: A33659

A;Molecule type: mRNA

A;Residues: 1-829 <SHI>

A;Cross-references: GB:X63629; NID:G35322; PIDN:CAA45177.1; PID:G35323

C;Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to

C;Genetics:

A;Gene: GDB:CDH3

A;Cross-references: GDB:132860; OMIM:114021

A;Map position: 16q24.1-16qter

C;Superfamily: cadherin; cadherin repeat homology

C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane prot

F;1-26/Domain: signal sequence #status predicted <SIG>

F;27-107/Domain: propeptide #status predicted <PRO>

F;108-829/Product: P-cadherin #status predicted <MAT>

F;110-215/Domain: extracellular #status predicted <EXT>

F;185-190/Region: cadherin repeat homology <CR1>

F;218-328/Domain: cadherin binding #status predicted

F;331-440/Domain: cadherin repeat homology <CR2>

F;441-548/Domain: cadherin repeat homology <CR3>

F;549-652/Domain: cadherin repeat homology <CR4>

F;653-677/Domain: transmembrane #status predicted <TMM>

F;678-829/Domain: intracellular #status predicted <INT>

F;785-800/Region: serine-rich

F;200,566/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 4369; DB 1; Length 829;

Best Local Similarity 100.0%; Pred. No. 8.4e-271;

Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASLLLLQVCWLQCAASPCRAVREAEVTLLEAGAGQEPGQALCKVFMGCEG 60

Db 1 MGLPRGPLASLLLLQVCWLQCAASPCRAVREAEVTLLEAGAGQEPGQALCKVFMGCEG 60

QY 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVWAPISVPENG 120

Db 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVWAPISVPENG 120

QY 121 KGFPFQRLNQLSKNRDRTKIFYSITGPGADSPPGVFAVEKETGWLILNKLPLDREETA 180

Db 121 KGFPFQRLNQLSKNRDRTKIFYSITGPGADSPPGVFAVEKETGWLILNKLPLDREETA 180


```

462 QY YTAEDPDK-ENOKISYRILRDPAGLWAMDPSGQVAVGTLDREDEQFVRNNIYEVAVLA 520
Db YTAEDPDK-ENOKISYRILRDPAGLWAMDPSGQVAVGTLDREDEQFVRNNIYEVAVLA 520
361 Db YTAQDPDRAQQORIKYVNGSDPAGLWAVHPENGLITAREQLDRE-SPTFKNSTYMAVLLA 419
YTAQDPDRAQQORIKYVNGSDPAGLWAVHPENGLITAREQLDRE-SPTFKNSTYMAVLLA 419
521 QY MDNGSPPTTGTGTLTLTLIDVNDHGVPPEPRGITTNCOSPVHRVHLNITDKLSPTSFPQ 580
Db MDNGSPPTTGTGTLTLTLIDVNDHGVPPEPRGITTNCOSPVHRVHLNITDKLSPTSFPQ 580
420 Db VDDGLPPATGTGTLTLTLIDVNDHGPEPEPRDIVICNSPPQVLTITDRLPNTGTGPR 479
VDDGLPPATGTGTLTLTLIDVNDHGPEPEPRDIVICNSPPQVLTITDRLPNTGTGPR 479
581 QY AQLTDDSDIYWTAENVESGDTVWLSKXFLKQDITDVHLSLSDHGKKEQLTVIRATVDCD 640
Db AQLTDDSDIYWTAENVESGDTVWLSKXFLKQDITDVHLSLSDHGKKEQLTVIRATVDCD 640
480 Db AELSHSGSGDSWAVEVNGGDTVALWLTLPLEQNYSVYLRFLDRQGDQDVIRAQVDCD 539
AELSHSGSGDSWAVEVNGGDTVALWLTLPLEQNYSVYLRFLDRQGDQDVIRAQVDCD 539
641 QY HGHVETC---POFWKG-GFILPVLGAVLALLFLVLVILLVLRKKRKIKEPILLPEDDTRD 696
Db HGHVETC---POFWKG-GFILPVLGAVLALLFLVLVILLVLRKKRKIKEPILLPEDDTRD 696
540 QY QRVESCAQKPRVDTGVPIVLAVLAVLALLVLLVLLVRRRKVKVKEPLLPEDDTRD 599
QRVESCAQKPRVDTGVPIVLAVLAVLALLVLLVLLVRRRKVKVKEPLLPEDDTRD 599
697 QY NVFYVYEGSGEGEEDDYDITOLHGLEARPEWLRNDVAPTLITPTMYRPPRANPDEIGN 756
NVFYVYEGSGEGEEDDYDITOLHGLEARPEWLRNDVAPTLITPTMYRPPRANPDEIGN 756
600 Db NIFYVYEGSGEGEEDDYDLSQHLGLDARPEV-RNDVAPFLMAAPQYRPPANPDEIGN 658
NIFYVYEGSGEGEEDDYDLSQHLGLDARPEV-RNDVAPFLMAAPQYRPPANPDEIGN 658
757 QY FTIENLKAANDTPTAPPYDTLLVFDYEGSGDAASLSLTSSAQDQDYDLNEWGSRF 816
FTIENLKAANDTPTAPPYDTLLVFDYEGSGDAASLSLTSSAQDQDYDLNEWGSRF 816
659 Db FIDENLKAANDTPTAPPYDLSLVFDYEGGSEATSSLSLNSASDQDQDYDLNEWGNRF 718
FIDENLKAANDTPTAPPYDLSLVFDYEGGSEATSSLSLNSASDQDQDYDLNEWGNRF 718
QY 817 KKLADMYGGGEDD 829
Db KKLADMYGGGEDD 829
719 Db KKLAEVYGGGEDE 731
KKLAEVYGGGEDE 731
RESULT 4
IOHUCE
cathelin 1 precursor [validated] - human
N:Alternate names: ARC-1; cell CAM 120/80; E-cadherin; epithelial cadherin; L-CAM; uvomor-
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1993 #sequence revision 30-Jun-1993 #text change 08-Dec-2000
R:Accession: S37654; S31430; S05475; S31460; S06716; A57171; JG2230; I52294; I52704; S253
R:Bussemakers, M.J.G.; van Bokhoven, A.; Mees, S.G.M.; Kemler, R.; Schaalken, J.A.
Mol. Biol. Rep. 17, 123-128, 1993
A:Title: Molecular cloning and characterization of the human E-cadherin cDNA.
A:Reference number: S37654; MUID:93211394; PMID:8459805
A:Accession: S37654
A:Reference number:
A:Molecule type: mRNA
A:Residues: 1-882 <BUS>
A:Cross-references: ENML:Z13009; NID:g31072; PIDN:CAA78353.1; PID:g31073
R:Kemler, W.; Warda, A.; Oda, T.; Hirohashi, S.; Kemler, R.; Birchmeier, W.
submitted to the ENML Data Library, December 1992
A:Description: Sequence of human E-cadherin cDNA.
A:Reference number: S31430

```

A:Accession: S31430
A:Molecule type: mRNA
A:Residues: 1-542, 'F', 544-882 <KEL>
A:Cross-references: EMBL:Z18923; NID:g31074; PIDN:CAA79356.1; PID:g31075
R:Mansouri, A.; Spurr, N.; Goodfellow, P.N.; Kemler, R.
Differentiation 38, 67-71, 1988
A:Title: Characterization and chromosomal localization of the gene encoding the human cel
A:Reference number: S05475; MUID:890331725; PMID:3263290
A:Accession: S05475
A:Molecule type: mRNA
A:Residues: 157-311 <MAN>
A:Cross-references: EMBL:X12790
A:Note: nucleotide sequence is not complete
R:Fixen, U.H.
submitted to the EMBL Data Library, March 1990
A:Reference number: S31460
A:Accession: S31460
A:Molecule type: mRNA
A:Residues: 265-392 <FRI>
A:Cross-references: EMBL:X52279; NID:g28821; PIDN:CAA36522.1; PID:g28822
R:Wheelock, M.J.; Buck, C.A.; Bechtol, K.B.; Damsky, C.H.
J. Cell. Biochem. 34, 187-202, 1987
A:Title: Soluble 80-kd fragment of cell-CAM 120/80 disrupts cell-cell adhesion.
A:Reference number: S06716; MUID:87280410; PMID:3611200

RESULT 5

A:Accession: S03160
 A:Molecule type: mRNA
 A:Residues: 157-884 <RIN>
 A:Cross-references: EMBL:X06339
 A:Note: part of this sequence, including the amino end of the mature protein, was confirmed by Behrens, J.; Loewrick, O.; Klein-Hitpaas, L.; Birchmeier, W.
 Proc. Natl. Acad. Sci. U.S.A. 88, 11495-11499, 1991
 A:Title: The E-cadherin promoter: Functional analysis of a G-C-rich region and an epithelial reference number: 149565; MUID:92107977; PMID:11763063
 A:Accession: I49565
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-15 <RES>
 A:Cross-references: GB:M81449; NID:G192325; PIDN:AAA37352.1; PID:G192326
 R:Tong, K.I.; Yau, P.; Overduin, M.; Bagby, S.; Porumb, T.; Takeichi, M.; Ikura, M.
 FEBS Lett 352, 318-322, 1994
 A:Title: Purification and spectroscopic characterization of a recombinant amino-terminal E-cadherin
 A:Reference number: S48735; MUID:95010732; PMID:7925993
 A:Accession: S48735
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 156-300 <TON>
 C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in many other cellular processes.
 C:Genetics:
 A:Gene: E-cadherin
 C:Superfamily: cadherin; cadherin repeat homology
 C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein
 F:1-27/Domain: signal sequence #status predicted <SIG>
 F:28-156/Domain: propeptide #status predicted <PRO>
 F:157-884/Product: E-cadherin, epithelial #status experimental <MAT>
 F:157-699/Domain: extracellular #status predicted <EXT>
 F:159-264/Domain: cadherin repeat homology <CR1>
 F:234-239/Region: cadherin binding #status predicted
 F:267-377/Domain: cadherin repeat homology <CR2>
 F:380-488/Domain: cadherin repeat homology <CR3>
 F:489-597/Domain: cadherin repeat homology <CR4>
 F:598-702/Domain: cadherin repeat homology <CR5>
 F:702-733/Domain: transmembrane #status predicted <TM>
 F:734-884/Domain: intracellular #status predicted <INT>
 F:842-855/Region: serine-rich
 F:560,639/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 54.78; Score 2388; DB 1; Length 884;
 Best Local Similarity 53.6%; Pred. No. 2.4e-444; Indels 66; Gaps 13;
 Matches 472; Conservative 123; Mismatches 222; Indels 66; Gaps 13;

QY 9 ASLLQLQV-CWLOCAASEP--CRAVFEAEVTLGAGAEQEPGQALGV-FMGCPCQPEA 64
 DB 10 ALLLLQLQVSSWL-CQELPESCSFGFSSEVYTFVPEHLEHGHVLRVFECTGRPT 68
 QY 65 LFSTNDNDFTVRNGETVQERRSLK----- 88
 DB 69 AFFSEDSRFKVTADGTTITVKRHLKHLKLETSFLVRARDSSHRELSTKVTLSKMGHHHRH 128
 QY 89 -----ERNP-LKIPPSKRI-LRRHKRDWVAPISVPENKGGPPQRLNOLSKNKDRDT 139
 DB 129 HHRDPASENPPELLMFPSVYVGLRQRKDWIIPISCPENKGGPPQRLNOLSKNKDRDT 188
 QY 140 KIFYSITGPGADSPPEGVFAVEKETGMLLNKLPDLREIAKYELFGHVSNGASVEDPM 199
 DB 189 KVFYSITGQADKPPVGVFTIERETGWLKVTQPLDREALAKYILYSHAVSSNGAEVDEM 248
 QY 200 NISITVDQNDHKPKFTQDFRGSVLEGLVPGTSMQVATDDEDAIYNGVAVSIHS 259
 DB 249 EIVITVDQNDNRPEFTQEVFEGSVAEGAVPGTSMQVATDDEDAIYNGVAVSIHS 308
 QY 260 QEPKDPHDLMTIHRSTGTTISVSSGLDREKVPETITICATDMDGSGSTTAVAVVEIL 319
 DB 309 QDPPELPHKMTVNRDGTGLSVLTSLGDRSYFTYTLVQADLQEGSLSTTAKAVITVK 368
 QY 320 DANDNAPMPQKYEAEHVPENAVCHEVQRLITVTDLADPNSPAWRAYILMGDGDGHPRT 379
 DB 369 DINDNAPVFNSTYQGVQVPEVNEVNIATLKVTDADDAPNTPAMKAVYTVV-NDPQQQFVV 427

RESULT 6

S34438

uvomorulin - mouse

C:Species: Mus musculus (house mouse)

C:Date: 22-Nov-1993 #sequence_revision 03-Nov-1995 #text_change 23-May-1997

C:Accession: S34438

R:Ringwald, M.; Baribault, H.; Schmidt, C.; Kemler, R.

Nucleic Acids Res. 19, 6533-6539, 1991

A:Title: The structure of the gene coding for the mouse cell adhesion molecule uvomorulin

A:Reference number: S34438; MUID:92093614; PMID:1754391

A:Accession: S34438

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-884 <RIN>

A:Cross-references: EMBL:X60975

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1992

C:Genetics:

A:Introns: 16/3; 57/1; 131/3; 179/3; 231/3; 280/1; 338/3; 381/3; 442/3; 524/2; 573/1; 641

C:Superfamily: cadherin; cadherin repeat homology

F:380-488/Domain: cadherin repeat homology <CR3>

Query Match 54.5%; Score 2379; DB 2; Length 884;

Best Local Similarity 53.5%; Pred. No. 9.1e-144; Indels 66; Gaps 13;

Matches 471; Conservative 122; Mismatches 222; Indels 66; Gaps 13;

QY 9 ASLLQLQV-CWLOCAASEP--CRAVFEAEVTLGAGAEQEPGQALGV-FMGCPCQPEA 64
 DB 10 ALLLLQLQVSSWL-CQELPESCSFGFSSEVYTFVPEHLEHGHVLRVFECTGRPT 68
 QY 65 LFSTNDNDFTVRNGETVQERRSLK----- 88
 DB 69 AFFSEDSRFKVTADGTTITVKRHLKHLKLETSFLVRARDSSHRELSTKVTLSKMGHHHRH 128
 QY 89 -----ERNP-LKIPPSKRI-LRRHKRDWVAPISVPENKGGPPQRLNOLSKNKDRDT 139
 DB 129 HHRDPASENPPELLMFPSVYVGLRQRKDWIIPISCPENKGGPPQRLNOLSKNKDRDT 188
 QY 140 KIFYSITGPGADSPPEGVFAVEKETGMLLNKLPDLREIAKYELFGHVSNGASVEDPM 199

Db 189 KVFYSITGQADKPPVGVFIIERETGWLKVTPQLDREAIKAYILYSHAVSSNGEAVEDEPM 248
 QY F:44-170/Domain: signal sequence #status predicted <SIG>
 F:44-170/Domain: propeptide #status predicted <PRO>
 F:171-895/Product: EP-cadherin #status predicted <MAT>
 F:171-718/Domain: extracellular #status predicted <EXT>
 F:173-278/Domain: cadherin repeat homology <CR1>
 F:248-253/Region: cadherin binding #status predicted
 F:281-391/Domain: cadherin repeat homology <CR2>
 F:394-502/Domain: cadherin repeat homology <CR3>
 F:503-610/Domain: cadherin repeat homology <CR4>
 F:611-714/Domain: cadherin repeat homology <CR5>
 F:719-743/Domain: transmembrane #status predicted <TM>
 F:744-895/Domain: intracellular #status predicted <INT>
 F:852-865/Region: serine-rich
 F:440.696/Binding site: carbohydrate (Asn) (covalent) #status predicted
 Query Match 54.4%; Score 2375; DB 1; Length 895;
 Best Local Similarity 54.2%; Pred. No. 1.7e-143;
 Matches 471; Conservative 128; Mismatches 214; Indels 56; Gaps 9;
 QY 11 LLLQVCMWCAASEPCRAVFRAREVILEAGCAEQEQCALGV-EMGCPQEQEPALFSTD 69
 Db 32 LCLLQVPSINADVSGCKPGFSAEYIFSVNRRELERGRKLGKVFNSDCTTRKGLDVG 91
 QY 70 NDDFTV-----RNGETVQERRS 86
 Db 92 DSRFVLPDGTIVKRVHVKLHKDKFTISTWDARGIKHSTNIAVASKRHSRGEAHSRS 151
 QY 87 LKERNPLKIFPSKRI-LRRHKRDVWVAPISVPENKGGPFPORLNQLKSKNDKDTKIYSI 145
 Db 152 ---KLPLVITPETHETGLKRRKRDWIPPIKVSNERGPFKRLVQIKSKNDRFNKYISI 208
 QY 146 TGGGADSPGGEVFAVEKETGMLLNKPLDREIYAKYELFGHAVSENGASVEDPMNIIIV 205
 Db 209 TQGGADNPQGVFRIEWETGMLVTRPLDREYDKYVLSHAVSENGSPVEEPMEIINV 268
 QY 206 TDQNDHKPKFTQDTPRGSVLEGVLPSTVMQVATDEDDAIYTVNGVVAISHQEPKDP 265
 Db 269 IDQNDNRKFTQDVPFRGVRGVPGTQWMAVSATDEDDNIDSLNGVLSYILKQDPPEP 328
 QY 266 HDLMEIHRSTGISVSSGLDREKVPYTLTIQATMDGDSGTTTAVAVVEILDANDNA 325
 Db 329 IPNLFTINRETVISLIGTGLDREKFPYTLTVQATDLEGAGLSEVKAIITQIDANDNA 388
 QY 326 PMFDPQKTEAHVPENAVGHEVQRLTVDLADPNFSPAEATYLMGDDGDGHTTTTTPES 385
 Db 389 PIFDEKTYTALVPENIEIGFEVQRLSVTDLDPGTPAWQAVYKIR-VNEGGEFFNITDPES 447
 QY 386 NOGILITRKGLDPEAKNQHLYVVEVNEAPFVVKLPSTATVTVHVEDVNEAPVFPESK 445
 Db 448 NOGILITRKGLDPEAKNQHLYVVEVNEAPFVVKLPSTATVTVHVEDVNEAPVFPESK 507
 QY 446 VVEVEGPIPTGPPVCVYTAEDPKEN-QKISYRIILRDPAGWLAMPDPSQVTAVTGLDRE 504
 Db 508 RVDSSEDLRGEKIIISLVAQDPDKQIQKLSYFIGNDFARMITVKNKGIVTGNGLDRE 567
 QY 505 DEQFVNNIYVWVLAMNGSPPTTGTGLTLLTLDVNDHGVPEPPEPITTCNQSPVZHV 564
 Db 568 SE-YVKNNTYTVIMLVTDGVSIGTGLTLLHVLVDNDNGVPSPRFTVTCWQDQNPQV 626
 QY 565 LNTDKLSPTSPPQAOLTDSDIYVTAENVNEEDGTIVLSLKKFLKQDTYDVHLSLSDH 624
 Db 627 LTIADIPPNYIPKYSLSHGSDITWAKELDSKTSMLLSPQQLKKGDYSIYVLLSDA 686
 QY 625 GNKEQLTVIRATVCDCHGVETCPGFWKGGFTLP-----VLGAVLALLFLLVLLVLRKK 680
 Db 687 QNPPQLTVVNAVTCSEGEKAIKQEKLVGGFDLPILVILGSLVALLILFLLVLLVLRKK 746
 QY 681 RKIKPELLLPEDDTRNDVYVYEGEGGEDDYDTQLHRLGEARPEVVLVNDVAPTIIIP 740
 Db 747 KVKKEPLLPEDDTRNDVYVYEGEGGEDDYDTQLHRLGEARPEVVLVNDVAPTIIIP 805
 QY 741 TPMYPRPAPNDPEIGNFIENLKAANTDPTAPPYPTLLVDFYEGSGSAASLSLTSAS 800

RESULT 7

LUXLCP
 EP-cadherin precursor - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
 C:Accession: B43785; A60128
 R:Ginsberg, D.; DeSimone, D.; Geiger, B.
 Development 111, 315-325, 1991
 A:Title: Expression of a novel cadherin (EP-cadherin) in unfertilized eggs and early xenopus embryos.
 A:Reference number: A43785; MUID:91372132; PMID:1893866
 A:Accession: B43785
 A:Molecule type: mRNA
 A:Residues: 1-895 <GIN>
 A:Cross-references: GB:X63720; NID:G64681; PIDN:CAA45252.1; PID:G64682
 A:Note: it is uncertain whether Met-1 or Met-16 is the initiator
 R:Angres, B.; Mueller, A.H.J.; Kellermann, J.; Hausen, P.
 Development 111, 829-844, 1991
 A:Title: Differential expression of two cadherins in Xenopus laevis.
 A:Reference number: A60128; MUID:91347911; PMID:1879345
 A:Accession: A60128
 A:Molecule type: protein
 A:Residues: 171-177, 'I', 179-183, 'K', 185-189, 'XI' <ANG>
 A:Note: the material sequenced may have contained U-cadherin as well as E-cadherin
 C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in cell adhesion, cell migration, and cell division.
 C:Superfamily: cadherin; cadherin repeat homology
 C:Keywords: calcium binding; cell adhesion; duplication; embryo; glycoprotein; transmembrane

Db 806 APHYRPRFSNDEIGNFIDENLDAADNDPTAPPYDSLIVFDYEGSGSEASLSLINGSNS 865
 Qy 801 DODQDYDYLNEGSRFKLADMYGGEDD 829
 Db 866 NDEHYNYLSWNGSRFKLADMYGGDDDE 894

RESULT 8
 S43064
 C:Species: African clawed frog
 C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
 C:Accession: S43064; S43065
 R:Kuehl, M.
 submitted to the EMBL Data Library, March 1994
 A:Reference number: S43064
 A:Accession: S43064
 A:Molecule type: mRNA
 A:Residues: 1-305 <KUE>
 A:Cross-references: EMBL:X78546; NID:9468816; PIDN:CAA55292.1; PID:9468817
 F:Herzberg, F.; Wildermuth, V.; Wedlich, D.
 Mech. Dev. 35, 33-42, 1991
 A:Title: Expression of Xbcad, a novel cadherin, during oogenesis and early development
 A:Reference number: S43065; MUID:92062581; PMID:1840622
 A:Accession: S43065
 A:Molecule type: mRNA
 A:Residues: 'NSA', 462-697, 'Q', 699-807, 'A', 809-840, 'N', 842-877, 'N', 879-883, 'N', 885-902, 'H'
 A:Cross-references: EMBL:X78546
 C:Superfamily: cadherin; cadherin repeat homology
 C:Keywords: calcium binding; cell adhesion; glycoprotein; transmembrane protein
 F:181-286/Domain: cadherin repeat homology <CR1>
 F:289-399/Domain: cadherin repeat homology <CR2>
 F:402-510/Domain: cadherin repeat homology <CR3>
 F:511-618/Domain: cadherin repeat homology <CR4>
 F:619-722/Domain: cadherin repeat homology <CR5>

Query Match 54.3%; Score 2371; DB 2: Length 905;
 Best Local Similarity 54.1%; Pred. No. 3.1e-143;
 Matches 469; Conservative 125; Mismatches 222; Indels 51; Gaps 8;

Qy 11 LLLQVWLCQAASEPRAVFRAEVLEAGGAEQFQALGV-FMGCPGQPFALFSTD 69
 Db 39 LCLQVVPISNDVSGQPGFSSANYTFVNRRELERGRKGLVNLVCTTRKGLVDVG 98
 Qy 70 NDPTVRNGTV-----QERRSLK 88
 Db 99 DSFRVLPGDTLVKRVHVKHDKDTRTITWDARGIKHSTNIVSVNKHRSGEARS 158
 Qy 89 ERNPLKIPPSKRI-LRRHKRDWVAPISVPENGKPPQRLNOLKSNKORDTKIFYSIIG 147
 Db 159 SELPVLTFPEKHTGLKRRDWIPPIKVSENERGPPFKELVQIKENKLSKVFYSIIG 218
 Qy 148 PGADSPGCVFAVEKETGWLKLLKPLDRERIAKVELFGHAVSNGASVEDPMNISIIVTD 207
 Db 219 QGADTPPEGIFRIEKEGVMQVTRPLDRREYKYLHSHAVSNGASVEEPMETITVTD 278
 Qy 208 QNDHKPKFTQDTPRGSLVGLVPGTSVMQVATDEDDAIVYNGVAVYSIHSQPKDHPD 267
 Db 279 QDNRRKFTQVPRGSRVREGVGTQKVMVSATDDSDSLNGVYAIYILKQDPPEPTP 338
 Qy 268 LMTIHRSTGTSIVISGLDREKVPYTLTIQATMDGCGSTTTAVAVVEILLDANDNAPM 327
 Db 339 NLFTINRETGIVSLIGLDRKFPYTLTVQAADLDGAGLTAEGKAVIEITDANDNAP 398
 Qy 328 FDPQKYEAHPENAVGHEVQRLVTDLDPNSPAWRATVILINGDDGDHFTITTHPESNQ 387
 Db 399 FDKTYTALVPENEVEFEVQRLSVTDLDMPGTAHQAVYKIR-VNEGGFNITTPESNQ 457
 Qy 388 GILTTKGLDFAEKQHTLYVEVNEAPFVLKLPSTSTATVHVHEDVNEAPVFPVPSKV 447
 Db 458 GILTTAKGLDFEVKQVVIQITVENAVPFSVPLPTSTATVTVVEDVNEAPVFPVPSV 517
 Qy 448 EYQEGTGPVCTAEPDKEN-QKISYRIILRDPAGWLAMPDPSGQVATVGLDREDE 506

Db 518 DVSEDLTRGEKIVSLVAQPDKKQIQKLSYFINDFARWLTKNGKIVTGNGLDRESE 577
 Qy 507 QFVRNNIYEVMLAMNDGSPPTTGTCTLLTLIDVNDHGFVPEPRQITTCNQSFVRHVLN 566
 Db 578 -YVKNNTYIVMLVTDGVPVGTGTGLTLHLVLDINDNGFVSPRVFTWCDQNPQVLT 636
 Qy 567 ITDKLSBHTSPQAOITDDSDIYMTAEVNEBGDTVLSLKKFLKQDITDVHLSLDHGN 626
 Db 637 ITDADIPPNYFYSVLSHSELTKAEIDSGKTSNLSPTQOLKKGDYSIYVLLADAQA 696
 Qy 627 KEQLTVIRATVCDCHGVETCTCPQWKGGFILP----VLGAVLALLFLLLVLLLVKRRKK 682
 Db 697 NRQLTVNATVCICEGKAICQEKLVAGFDLPILVILGASILALLILSLLLFLKRRKV 756
 Qy 683 IKPELILLPEDDRDNRVYVYEGEGGEEDQDYITQLHRCLEARPEVVLNRDVAPTIIP 742
 Db 757 VKEPLLLPEDDRDNRNIFYEGEGGEEDQDYDLSQLHRLDARPD-IMENDVVPVLTMSVP 815
 Qy 743 MYRPRPANDEIGNFTIENLKAANTDPTAPPYDITLLVFDYEGSGDAASLSLTSSASDQ 802
 Db 816 HYRPRPSNDEIGNFTIDENLDAADNDPTAPPYDLSLVFDYEGSGSEASLSLNSNSNN 875
 Qy 803 DQDYVLENGSRFKLADMYGGEDD 829
 Db 876 EHDYNYLNDWGPFRKFLADMYGGDDDD 902

RESULT 9
 IJBOCP
 P-cadherin - bovine (fragment)
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
 C:Accession: S11694
 R:Liaw, C.W.; Cannon, C.; Power, M.D.; Kiboneka, P.K.; Rubin, L.L.
 EMBO J. 9, 2701-2708, 1990
 A:Title: Identification and cloning of two species of cadherins in bovine endothelial cell
 A:Reference number: S11693; MUID:90360979; PMID:2390969
 A:Accession: S11694
 A:Molecule type: mRNA
 A:Residues: 1-491 <LIA>
 A:Cross-references: EMBL:X53614; NID:6166; PIDN:CAA37676.1; PID:9833777
 C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to
 C:Superfamily: cadherin; cadherin repeat homology
 C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane prot
 F:1-491/Product: P-cadherin (fragment) #status predicted <NAT>
 F:1-314/Domain: extracellular (fragment) #status predicted <EXT>
 F:1-102/Domain: cadherin repeat homology (fragment) <CR3>
 F:103-210/Domain: cadherin repeat homology <CR4>
 F:211-314/Domain: cadherin repeat homology <CR5>
 F:315-339/Domain: transmembrane #status predicted <TMM>
 F:340-491/Domain: intracellular #status predicted <INT>
 F:447-462/Region: serine-rich
 F:228/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 53.2%; Score 2323; DB 1: Length 491;
 Best Local Similarity 88.4%; Pred. No. 1.4e-140;
 Matches 434; Conservative 23; Mismatches 34; Indels 0; Gaps 0;

Qy 339 ENAVGHEVQRLTVTDLDAPNSPAWRATVILINGDDGDHFTITTHPESNQGLITTKGLDF 398
 Db 1 ENTVSHEVQRLTVTDLDAPNSPAWRATVIRVGGDGDHFTITTPESNQGLITTKGLDF 60
 Qy 399 EAKQHTLYVEVNEAPFVLKLPSTSTATVHVHEDVNEAPVFPVPSKVVEQSGIPTGEP 458
 Db 61 EAKQHTLYVEVNEAPFVLKLPSTSTATVHVHEDVNEAPVFPVPSKVVEQSGIPTGEP 120
 Qy 459 VCVTAEPDKENQKISYRIILRDPAGWLAMPDPSGQVATVGLDREDEQFVRNNIYEV 518
 Db 121 ICAYTARPDGSKQKISVHILRDPAGWLAMPDPSGQVATVGLDREDEQFVRNNIYEV 180
 Qy 519 LAMNGSPPTTGTCTLLTLIDVNDHGFVPEPRQITTCNQSFVRHVLNITDKLSPTSP 578

Db 181 LATDDGSPPTGTTGTTLLTMDINDHGPVPEPRQITICNOSPVPQVLNITDKLSHTAP 240
 Qy 579 FOAQLTDDSDIYTAENEGDVTWLSLKKFLKQDYDVHLSLSDHGNKEQLTVIRATVC 638
 Db 241 FOAQLTHSDSVYTAENEGDAVALSLKKFLKQGEYDVHLSLSDHGNKEQLTVIRATVC 300
 Qy 639 DCHGVETCPGPKGKGFILVGLAVIALFLVLLVLLVLRKKIKIPELILLPEDDTRDNV 698
 Db 301 DCHGNMVTCDPTWGTWGLLILGALALLLLVLLVLRKKIKIPELILLPEDDTRDNV 360
 Qy 699 FYTGEEGGEGEDQDYDTQLHRLGLEARPEVLRNDVAPTIPTMYRPRPANDDEIGNFI 758
 Db 361 FYTGEEGGEGEDQDYDTQLHRLGLEARPEVLRNDVAPTIPTMYRPRPANDDEIGNFI 420
 Qy 759 IENLKAANTDPTAPPYDTLAVFDYEGSGDAASISSLTSSASDQDQDYDLNWSGRPKK 818
 Db 421 IENLKAANTDPTAPPYDLSLIVFDYEGSGDAASLSSLTSSDQDQDYDLNWSGRPKK 480
 Qy 819 LADMYGGGEGEDD 829
 Db 481 LADMYGGGEGEDD 491

RESULT 10
 IUCHCL
 B-cadherin precursor, hepatic - chicken
 N:Alternate names: L-CAM; liver cell adhesion molecule
 C:Species: Gallus gallus (chicken)
 C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 25-Oct-1996
 C:Accession: A30201; A29866; B29866
 R:Sorkin, B.C.; Hemperly, J.J.; Edelman, G.M.; Cunningham, B.A.
 Proc. Natl. Acad. Sci. U.S.A. 85, 7617-7621, 1988
 A:Title: Structure of the gene for the liver cell adhesion molecule, L-CAM.
 A:Reference number: A30201; MUID:89017248; PMID:3174655
 A:Accession: A30201
 A:Molecule type: mRNA
 A:Residues: 1-81 <SOR>
 A:Cross-references: EMBL:J04074
 R:Gallin, W.J.; Sorkin, B.C.; Edelman, G.M.; Cunningham, B.A.
 Proc. Natl. Acad. Sci. U.S.A. 84, 2808-2812, 1987
 A:Title: Sequence analysis of a cDNA clone encoding the liver cell adhesion molecule, L-
 A:Reference number: A29866; MUID:87204217; PMID:3472238
 A:Accession: A29866
 A:Molecule type: mRNA
 A:Residues: 51-887 <GAI>
 A:Cross-references: EMBL:M16260
 A:Accession: B29866
 A:Molecule type: protein
 A:Residues: 161-172;323-336;386-407;533-551 <GA2>
 C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought
 C:Genetics:
 A:Introns: 23/3; 62/1; 138/1 183/3; 235/3; 284/1; 342/3; 385/3; 447/3; 529/2; 577/1; 652/1
 C:Superfamily: cadherin; cadherin repeat homology
 C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; liver; transmembr
 F:1-26/Domain: signal sequence #status predicted <SIG>
 F:27-160/Domain: propeptide #status predicted <PRO>
 F:161-887/Product: E-cadherin, hepatic #status experimental <MAT>
 F:161-887/Product: E-cadherin, hepatic #status predicted <EXT>
 F:161-704/Domain: extracellular #status predicted <CR1>
 F:163-268/Domain: cadherin repeat homology <CR1>
 F:238-243/Region: cadherin binding #status predicted
 F:271-381/Domain: cadherin repeat homology <CR2>
 F:384-493/Domain: cadherin repeat homology <CR3>
 F:494-601/Domain: cadherin repeat homology <CR4>
 F:602-704/Domain: cadherin repeat homology <CR5>
 F:736-887/Domain: transmembrane #status predicted <TM>
 F:844-857/Region: intracellular #status predicted <INT>
 F:291,346,564,643/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 52.0%; Score 2274; DB 1; Length 887;
 -Best Local Similarity 52.7%; Pred. No. 4,6e-137;
 Matches 461; Conservative 122; Mismatches 236; Indels 55; Gaps 13;

Qy 8 LASLILLQVCMWLQCAASEPCRAVFEAEVLEAGABQEPGQALGV-FMGCQGPALF 66
 Db 16 LVLLILLQVCMWLQCAASEPCRAVFEAEVLEAGABQEPGQALGV-FMGCQGPALF 75
 Qy 67 -STD-----ND-----FTVANGETVQERRS----- 86
 Db 76 VPTDTRFKVNGDGVVSKRPLTLYGRKISFTIYAQDANGKHSARVTVGHRHRRHHNH 135
 Qy 87 -LKERNPLKI-FP--SKRIILRRHRDWWVAPISVPENKGFPPQRLNQLKSNKRDRTKIF 142
 Db 136 HLQDTTAVLTFFKHPDGLRRQRQSDVPIPSICLENHGRGPYPMRLVQIKSNKDKESKV 195
 Qy 143 YSITGPGADSPGEGVFAVEKETGWLKLNKPLDREIAKYLFGHAVSENGASVEDPNIS 202
 Db 196 YSITGPGADSPGEGVFAVEKETGWLKLNKPLDREIAKYLFGHAVSENGASVEDPNIS 255
 Qy 203 IIVTDQNDHKFKFTQDTRFGSVLEGVLPFTSVMOVTATDEDDAIYTYNGVVAISIHSQEP 262
 Db 256 ITVMDQNDKVPFIKEVFGVIEENAKPGISVMTVNATDADDAVNTDNGIVSVISQOP 315
 Qy 263 KDPHDLMTIHRSTGTSIVSSGLDREKVPETITIOATMDGSGSTTAVAVVEILDAN 322
 Db 316 PRPHQMFETIDPAKGIISVLGTLDRETTNYTLIVQATDQEGKGLSNTAIIIEVDAN 375
 Qy 323 DNAPMFDQKYEAVHPENAVGHEVQRLVTITDLPNGSPAVRATYILMGDGDGDFHTTTH 382
 Db 376 DNIPIFNMTYEGVVVENKPGTEVARLTVTDQDAPGSPAWQAVYHIKSGNLDAFSITD 435
 Qy 383 PESNGILTTTRKGLDFAKNOHTLYVEVNEAPVFLKLPSTATIVVHVEDVNEAPVFP 442
 Db 436 PSTNNGILTKAKGLDYETKSYDLVVTVENKVPISVPIITLSTASVLTVLDWNEPVPFP 495
 Qy 443 PSKVVEVQEGIPTEPCVCTAEPDKE-NQKISVRIILDPAGLAMPDPSGQVTAAGTL 501
 Db 496 PIKRVGVPEDLFGVQGVTSYTAQDPDRMRQKITRMGSDDPAGWLYHPENGIIVTAQPL 555
 Qy 502 DREDEQFVANNIYEVWVLAMNDGSPPTTGTGLTLLTIDVNDHGVPEPRQITICNOSP 561
 Db 556 DRESVHAI-NSTYKAIILAVDNGIPDTTGTGLTLLQDNDVNDGPTPEPRSFICSRQPE 614
 Qy 562 RHVLNITDKLSPTSPQAQLTDDSDIYTAENEGDVTWLSLKKFLKQDYDVHLSL 621
 Db 615 KQILSIVDKLPPHTYTPFKAALHSGSSNNWTVIRGQ-DELAAGLKKLECEGFNIFVKL 673
 Qy 622 SDHGNKEQLTVIRATVCCHGHVETCP--GPWKGGF----ILPVLGAVIALFLLLVLL 675
 Db 674 TDSQGAQVTVKAQVCEGCTAKNCERRSYVGLGVPAILGILGTLALLILLILL 733
 Qy 676 LVKRRKIKIPELILLPEDDTRDNVFFYEGGEGEDQDYDTQLHRLGLEARPEVLRNDVA 735
 Db 734 FARRKKEKEPILLPEDDTRDNVFFYEGGEGEDQDYDTQLHRLGLEARPEVLRNDVA 792
 Qy 736 PIITPTMYRPRPANDDEIGNFIENLKAANTDPTAPPYDLSLIVFDYEGSGDAASLSSL 795
 Db 793 PPLMAQPRPRPANDDEIGNFIENLKAANTDPTAPPYDLSLIVFDYEGSGDAASLSSL 852
 Qy 796 TSSASDQDQDYDLNWSGRPKKCLADMYGGGEGEDD 829
 Db 853 NSSASDQDQDYDLNWSGRPKKCLADMYGGGEGEDD 886

RESULT 11

S47518
 cadherin - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 01-Feb-1995 #sequence_revision 10-Nov-1995 #text_change 17-Mar-1999
 C:Accession: S47518
 R:Tool, O.; Fujii, G.; Tashiro, K.; Shiohara, K.
 Biochim. Biophys. Acta 1219, 121-128, 1994
 A:Title: Molecular cloning of cDNA for XTCD-1, a novel Xenopus cadherin, and its express
 A:Reference number: S47518; MUID:94368839; PMID:8086449
 A:Accession: S47518
 A:Molecule type: mRNA

A;Residues: 1-871 <too>
C;Superfamily: cadherin; cadherin repeat homology
F;151-256/Domain: cadherin repeat homology <CR1>

Query Match
Best Local Similarity 50.2%; Score 2192.5; DB 2; Length 871;
Matches 460; Conservative 126; Mismatches 227; Indels 73; Gaps 16;

QY 1 MGLPR-----GFLASLLLOVC-----WLQCAASEPCRAVFPREAEVTLLEAGAGAEQPGQAL 51
DB 1 MGLRPMLLGAVLLTLITQVGLAEWTQ-----CRNGFSKEKYSFLV-PKNLETDKAL 53
QY 52 GKV-FMGCPGQEPALFSTNDNDFVRNGETVQERRSLKERNPLKIF----- 96
DB 54 GRVFNCEGVRVQFASKDPNFHKDGTVIYIKNPAKMDNRKTRFVLAWETKGVYST 113
QY 97 -----PSKILRRHRKRDWVAVPISVPENKGGPFPORLNOLK 132
DB 114 NITLKREGRHRQDLFGSKCHSHPKSEGLKQKRDWVIPPILVSENEKGGPFPKRIQIK 173
QY 133 SNKDRDKIFYSITPGADSPPEGVFAVEKETGMLLNKPLDRBEIAKYELFHAHSVNG 192
DB 174 SSYAKEVKVYSITGQADTFPEGVFAIGREDGWLNVTRPLDREADNYYLFFSHAVSSNG 233
QY 193 ASVEDPMNISIIITDQNDHKPKFTQDTFRGSLVGLPGLTSMQVATDDEDAIYTVNGV 252
DB 234 ANVEDPMELIKVQDQNDNDVFTQSVFEGSVPEGKGTAVMTVSATDADSDVDMXNGV 293
QY 253 VAYSIHQEPKDPHLMFTTHRSRTGTSVSSGLDRKVPYEYTLTIQATMD--GDGSTTT 311
DB 294 ITYSILNQEPKEPTNKNFTTHSESGLISVLTTGLDREKNPVYTLTIQAADGEGKDRIT 353
QY 312 AVAVVEILDANDAMPDPQKYEAHVPENAVGHVQRLTWTDDAPNSPAWRATYILMG 371
DB 354 ATALIVMDNDNPVDFPTQYAKVENEVGHVQRLTWTDEIGTDANNAVYKILK 413
QY 372 DGGHFTTTHPESNOGLITRKLDFEAKNQHTLYVEVNEAPFVLKLTSTATIVHY 431
DB 414 NEANYFSIQT-DTGNIGLLTKVGLDYELKQYLSLVITNKANFSVPLQTSTATVSV 472
QY 432 EDVNEAPVFPVPPSKVVEQGIPTGEPVVCVYTAEDPKK-NOKISYELLDPAGWLAMP 490
DB 473 EDVNEAPFIFPPVKEVSSDLPSGQVAVTYTAQDPKQNKQKITVYIGNDPGRWSVAK 532
QY 491 DSGQVAVGLDRDEQFVNNIYVVMVLAMNGSPPTGTGTLTLLTLDVNDHGPVPEP 550
DB 533 DNGIVTGNGLDRE-SKFLVNNYKVILLAADSGSPSATGTGLVNLNLDVNDNGFLEP 591
QY 551 RQITICNQSPVRHVLNTDKDLSPTSPFOQLTDDSDIYVTAEVNEEGTVVLSL--KK 608
DB 592 QSESFCQKQCFRFTIIDRLSPNTYFYKAELTGESNENWTAIVTDK----ILELRPKK 647
QY 609 FLQDQYDVHLSLSHGNKQLTVIRATVCDCHGVETC--PGPWKGGF-----ILPVLGA 662
DB 648 ELBIGQYDVMITLSDSGLSNVTKLHTICQGDGRKQCEKAAIAGGLGISAVILGG 707
QY 663 VLALLFLLVLLVLRKKRIKPLLPEDDTRDNVYFEGEGGEDQDYITQLHRLG 722
DB 708 ILALLLLLLLLLVRKKVVKPEPLPPEDETRDNVFSDEEGGEDQDFDLSQLHRLG 767
QY 723 EAPPEVLRNDVAPTIPTPWRPRPANDEIGNFIENLKAANTDPTAPPDYTLVFPY 782
DB 768 DAPDPVT-RNDVAP-VLAAQYPRPANDEIGNFIENLKAANTDPTAPPDYTLVFPY 825
QY 783 EGSQSDAASLSSTSSASDQDQDYVILNEMGSRFKKLADMYGGED 828
DB 826 EGSQSEAASSLSSNSNSDLQDYSALNDNGPFTKLADMYGGED 871

RESULT 12
1JHUCN
cadherin 2 precursor - human
N;Alternate names: N-cadherin; neuronal cadherin

C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A38870; S11487; JQ0751; S13799
R;Reid, R.A.
Submitted to the EMBL Data Library, November 1990
A;Reference number: A38870
A;Accession: A38870
A;Molecule type: mRNA
A;Residues: 1-906 <RE1>
A;Cross-references: EMBL:X54315; NID:G34998; PIDN:CAA38213.1; PID:G34999
R;Reid, R.A.; Hemperly, J.J., 1990
Nucleic Acids Res. 18, 5896, 1990
A;Title: Human N-cadherin: nucleotide and deduced amino acid sequence.
A;Reference number: S11487; MUID:91016946; PMID:2216790
A;Accession: S11487
A;Molecule type: mRNA
A;Residues: 1-340, 'N', 342-698, 'R', 700-704, 'F', 706-906 <RE2>
A;Cross-references: EMBL:X54315
A;Note: this sequence has been revised in reference A38870
R;Walsh, F.S.; Barton, C.H.; Putt, W.; Moore, S.E.; Kelsell, D.; Spurr, N.; Goodfellow, I.
J. Neurochem. 55, 805-812, 1990
A;Title: N-cadherin gene maps to human chromosome 18 and is not linked to the B-cadherin
A;Reference number: JQ0751; MUID:90347462; PMID:2384753
A;Accession: JQ0751
A;Molecule type: mRNA
A;Residues: 160-194, 'IR', 197-211, 'L', 213-227, 'Q', 229, 'N', 231-235, 'G', 237-248, 'T', 250-356,
A;Cross-references: GB:M34064
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion and are thought to
C;Genetics:
A;Gene: GDB:CDH2; NCAD
A;Cross-references: GDB:128185; OMIM:114020
A;Map position: 18q12.1-18q12.1
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane prot
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-159/Domain: propeptide #status predicted <PRO>
F;160-906/Product: N-cadherin #status predicted <NCA>
F;160-714/Domain: extracellular #status predicted <EXT>
F;162-267/Domain: cadherin repeat homology <CR1>
F;237-242/Region: cadherin binding #status predicted
F;270-382/Domain: cadherin repeat homology <CR2>
F;385-497/Domain: cadherin repeat homology <CR3>
F;500-605/Domain: cadherin repeat homology <CR4>
F;606-712/Domain: cadherin repeat homology <CR5>
F;715-746/Domain: transmembrane #status predicted <TM>
F;747-906/Domain: intracellular #status predicted <INT>
F;865-878/Region: serine-rich
F;190,273,325,402,572,622,651,692/Binding site: carbohydrate (Asn) (covalent) #status pre

Query Match 39.1%; Score 1707; DB 1; Length 906;
Best Local Similarity 42.7%; Pred. No. 7,3e-101;
Matches 390; Conservative 129; Mismatches 292; Indels 102; Gaps 27;

QY 6 GPLASLLLOVCWLOCAASEP-----CRAVFPREAEVTLLEAGAGAEQPGQALGKV-FMGCP 59
DB 6 GALTLLPALLLQALQVSAEAGEALCKTGPFPE-DVYSAVLKQVHGGQPLINVKFSNCN 64
QY 60 GQ-----BPALFSTNDND--FTVRN-----GETVOER----- 84
DB 65 GKRKQVYESSPEADFKVDEGDMVAVRSFPLSSSEHAKFLIYAQDKETQKQVAVKLSLK 124
QY 85 -----RSUKERNPLK--IFP---SKRI--LRHRKRDWVAVPISVPENKGGPFPQRLNOLK 132
DB 125 PTLTEESVKEGAEEVIEVFPQFSKSHQKQKRDWVIPPILVSENEKGGPFPQELVIR 184
QY 133 SNKDRDKIFYSITPGADSPPEGVFAVEKETGMLLNKPLDRBEIAKYELFHAHSVNG 192
DB 185 SDRKLVRLYTVTGPAGDQPTGIFINPISGQSLVTKPLDRQIARFHURAHAVDNG 244
QY 193 ASVEDPMNISIIITDQNDHKPKFTQDTFRGSLVGLPGLTSMQVATDDEDAIYTVNGV 252
DB 245 NQVENPIDIVINVIDMNDNRPEFLHQVWNGVTPGSKGTVMYTVMTADDDP-NALNGM 303

253 VAYSIHQEPKDPHDLMTTHRSFTGTSIVSSGLDREKVEYTLTQATDMOG---DGSST 309
 304 LRVIVSCAPSTSPNNFTINNETGDIITVAAGLDREKVKQYTLTQATDMGEPYGLS 363
 310 TTAVALVEILLDANDANAPFQPKYEAHVHPENAVGHEVQRLTVDLADPNSPAWRATYLM 369
 364 NTATAVITVDVNDNPPFTAMTFYGVGVRNVDIIIVANLTVTDKQPHTPANNAYRIS 423
 370 GGGDGDHFTITTHPESQGLITTRKGLDPEAKQHTLYVEVNEAPVLKL---PTSTAT 426
 424 GGDPTGFAICTDPSNDGLVTVPKPIDFTETNMFVLTAAENQVPLAKGIQHPQSTAT 483
 427 IVVHVEDVNEAPVFPVPSKVEVQEGPTGTCVYTAEDPK-ENOKISYRILRDPAGW 485
 484 VSVTVIDVNEPYPAPNPKIIRQEGHAGTMTLFTTAQDPDRYMQNIRYTKLSDPANW 543
 486 LAMPDPSQVTAAGTLDREDEBQFVRNNIYEVVLMADNGSPPTTGTGTLTLLTLDVNDHG 545
 544 LKIDPVNGQITTIIVLDRESN-VKNNIYNATFLASDNGIPWPGTGTGLQIYLLDINDNA 602
 546 PVPSPRQITICNOSPVRHVNIT--DKDLSPTSPFOAQLTDDSDIY---WTAEVNEEGD 600
 603 PQVLQPEAETC-ETPDPSNINITALDYIDFNAGPFAFLPLGFTVTKRNTI-TRLNGD 660
 601 TVVLSLK-KFLKQDTYVHLSLSDHGN--KEQLTVIRATVCDCHG-----VETCPGPKW 652
 661 FAQLNKKLFLEAGIYEVPIITDSGNPKNSISLAVKVCQDSNGDCTDVRIVGAGL 720
 653 GGFILPVLGAVLALLFLLVLLLV-----RKRKIKKPELLLPEDDTRDNDVYFG 702
 721 G-----TGAIITALLCIILLLVLMFVVMKRDKERQAKQLLIDPEDDVRDNLKYD 774
 703 BEGGGEEDDYDITQLHGLEAREV-----VLNDVAPITPIPMYRPPA--NPDEIG 755
 775 BEGGGEEDDYDLQLOQPTVEPDAIKPVGIRWDERP-IHAEPQYVRSAAHPFGDIG 833
 756 NFITENKAANTDPTAPPYDITLLVFDYEGSGSDAASLSLTSSASDQDQDYDYLNEGSR 815
 834 DFINEGLKAANDPTAPPYDITLLVFDYEGSGSTAGLSLSSSSSGEGEQDYDYLNDWGR 893
 816 FKKLADMYGGGED 828
 894 FKKLADMYGGGDD 906

RESULT 13
 IUCHGN
 N-cadherin precursor, neuronal - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
 C:Accession: A29964
 R:Hatta, K.; Nose, A.; Nagafuchi, A.; Takeichi, M.
 J. Cell Biol. 106, 873-881, 1988
 A:Title: Cloning and expression of cDNA encoding a neural calcium-dependent cell adhesion molecule
 A:Reference number: A29964; PMID:86153917; PMID:2831236
 A:Accession: A29964
 A:Molecule type: mRNA
 A:Residues: 1-912 <HAT>
 A:Cross-references: GB:X07277; NID:963649; PIDN:CAA30258.1; PID:G63650
 C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in a variety of cellular processes including cell adhesion, cell migration, and cell differentiation.
 C:Superfamily: cadherin; cadherin repeat homology
 C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein
 F:1-28/Domain: signal sequence #status predicted <SIG>
 F:29-164/Domain: propeptide #status predicted <PRO>
 F:165-912/Product: N-cadherin #status predicted <MAT>
 F:165-720/Domain: extracellular #status predicted <EXT>
 F:167-272/Domain: cadherin repeat homology <CR1>
 F:242-247/Region: cadherin binding #status predicted
 F:275-387/Domain: cadherin repeat homology <CR2>
 F:390-502/Domain: cadherin repeat homology <CR3>
 F:505-611/Domain: cadherin repeat homology <CR4>
 F:612-720/Domain: cadherin repeat homology <CR5>
 F:721-752/Domain: transmembrane #status predicted <TM>

F:753-912/Domain: intracellular #status predicted <INT>
 F:869-864/Region: serine-rich
 F:278,330,407,578,628,657/Binding site: carbohydrate (Asn) (covalent) #status predicted
 Query Match 39.0%; Score 1704.5; DB 1; Length 912;
 Best Local Similarity 42.6%; Pred. No. 1.1e-100;
 Matches 390; Conservative 132; Mismatches 284; Indels 109; Gaps 27;

QY 7 PLASILL--LQVWLOCASEP-CRAVPR-...AETLEAGAGAEQPGQALGV-FWGC 58
 DB 14 PLALMLAALQAPKATCEDMLCKMGFPEDVHSAVRSVHG-----GQPLLNRVQSC 68
 QY 59 PQCEPALF-STDNDDFTV-----ENGETVQERR----- 85
 DB 69 DENRKIYFGSSEPEDFRVGEDGVVVAERSFOLSABPTFVVVSARDKETQEBWQVKVLT 128
 QY 86 -----SLKERNPLK--IPFSKRI-----LRRHKRWVYAPISVPENKGGPPQRLNOL 131
 DB 129 EPAFTGASEKQDKKIEDIIFPQQYKQSSHLKROKRWIPPIINLPENSRGFFPQELVRI 188
 QY 132 KSKNDKDTKI FYSITGPGADSPPGVFAVEKETGMLINKPLDREELAKYELFGHAYSEN 191
 DB 189 RSDRXSLSLAYSVTGPADQPTGIFINPISGLSVTKELDREQIASFLRAHADV 248
 QY 192 GASVEDPMNISIIITDQNDHKPKFTQDTRGVSLEGLVPGTSVMQVTTATDEDDAIYNG 251
 DB 249 GNQVENPIDIVINVIDMNDNRPEFLHQVWNGTVPEGSKPGYVMTVTAIDADDP-NAQNG 307
 QY 252 VVAYSIHQEPKDPHDLMTTHRSFTGTSIVSSGLDREKVEYTLTQATDMOG---GS 308
 DB 308 MLRYKILQASGSPSPNNFTINNETGDIITVAAGLDREKVKQYTLTQATDMGEPYGL 367
 QY 309 TTTAVAVVEILLDANDANAPFQPKYEAHVHPENAVGHEVQRLTVDLADPNSPAWRATYLI 368
 DB 368 SNTATAVITVDVNDNPPFTAMTFYGVGVRNVDIIIVANLTVTDKQPHTPANNAYQM 427
 QY 369 MCGDGDHFTITTHPESQGLITTRKGLDPEAKQHTLYVEVNEAPVLKL---PTSTA 425
 DB 428 TCGDPTGQITLTDPSNDGLVTVPKPIDFTETNMFVLTAAENQVPLAKGIQHPQSTA 487
 QY 426 TTVHVEDVNEAPVFPVPSKVEVQEGPTGTCVYTAEDPKENQKIS--YKILRPPA 483
 DB 488 TVSIITVIDVNESPYFVPNPKLVRQEGGLAGSMLTFTTARDPDRYMQTSRYSKLSDDPA 547
 QY 484 GWLAMPDPSQVTAAGTLDREDEBQFVRNNIYEVVLMADNGSPPTTGTGTLTLLTLDVND 543
 DB 548 NWLKIDPVNGQITTTAVLDRE--SIYVQNNMYNATFLASDNGIPWPGTGTGLQIYLLDIND 606
 QY 544 HGPVPEPRQITICNOSPVRHVNIT--DKDLSPTSPFOAQLTDDSDIY---WTAEVNEE 598
 DB 607 NAFQVNPKEATTC-ETLQPNANITAVDPDIDPNAGPFAFLPSPSPISKRNTI-VRI 664
 QY 599 GDTVVLSLK-KFLKQDTYVHLSLSDHGN--KEQLTVIRATVCDCH-----GHVETCPGP 650
 DB 665 GCHAQLSLRIFLEAGIYDVPITVDSGNPHASSTSVLKVKVCQCDINGDCTDVRIVGA 724
 QY 651 WKGGFILPVLGAVLALLFLLVLLLV-----RKRKIKKPELLLPEDDTRDNDVY 700
 DB 725 GUG-----TGAIITALLCIILLLVLMFVVMKRDKERQAKQLLIDPEDDVRDNLK 778
 QY 701 YGEEGGEEDDYDITQLHGLEAREV-----VLNDVAPITPIPMYRPPA--NPDE 753
 DB 779 YDEEGGEEDDYDLSQLQOPDTVEPDAIKPVGIRWDERP-IHAEPQYVRSAAHPFGD 837
 QY 754 IGNFIENLKAANTDPTAPPYDITLLVFDYEGSGSDAASLSLTSSASDQDQDYDYLNEGW 813
 DB 838 IGDFINEGLKAANDPTAPPYDITLLVFDYEGSGSTAGLSLSSSSSGEGEQDYDYLNDWG 897
 QY 814 SRFKKLADMYGGGED 828
 DB 898 PRFKKLADMYGGGDD 912

RESULT 14
IUMSCN
N-cadherin precursor, neuronal - mouse
C:Species: Mus musculus (house mouse)
C>Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C/Accession: A32759; #Accession: A32759
R: Miyatani, S.; Shimamura, K.; Hatta, M.; Nagafuchi, A.; Nose, A.; Matsunaga, M.; Hatta, S.
Science 245, 631-635, 1999
A>Title: Neural cadherin: role in selective cell-cell adhesion.
A/Reference number: A32759; MUID: 83346748; PMID: 2762814
A/Accession: A32759
A/Molecule type: mRNA
A/Residues: 1-906 <MI>
A/Cross-references: GB:M31131; NID: g192327; PIDN: AAA37353.1; PID: g309125
R: Miyatani, S.; Copeland, N.G.; Gilbert, D.J.; Jenkins, N.A.; Takeichi, M.
Proc. Natl. Acad. Sci. U.S.A. 89, 8443-8447, 1992
A>Title: Genomic structure and chromosomal mapping of the mouse N-cadherin gene.
A/Reference number: A46163; MUID: 92409532; PMID: 11528849
A/Accession: A46163
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 839-906 <MI2>
A/Cross-references: GB:S45011; NID: g256010; PIDN: AAB23356.1; PID: g256011
A/Note: sequence extracted from NCBI backbone (NCBIN:113759, NCBI:113760)
C/Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in cell-cell adhesion.
C/Superfamily: cadherin; cadherin repeat homology
C/Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-159/Domain: propeptide #status predicted <PRO>
F:160-906/Product: N-cadherin #status predicted <MAT>
F:160-714/Domain: extracellular #status predicted <EXT>
F:162-267/Domain: cadherin repeat homology <CR1>
F:237-242/Region: cadherin binding #status predicted
F:270-382/Domain: cadherin repeat homology <CR2>
F:385-497/Domain: cadherin repeat homology <CR3>
F:500-605/Domain: cadherin repeat homology <CR4>
F:606-714/Domain: cadherin repeat homology <CR5>
F:715-746/Domain: transmembrane #status predicted <TM>
F:747-906/Domain: intracellular #status predicted <INT>
F:865-878/Region: serine-rich
F:190,273,325,402,572,651,692/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 38.9%; Score 1697.5; DB 1; Length 906;
Best Local Similarity 42.8%; Pred. No. 2.9e-100;
Matches 390; Conservative 125; Mismatches 236; Indels 101; Gaps 27;

Qy 5 RGPLASLL--LLQVWVLCQAASEPCRAVFEAEVTLLEAGAEPPGQALQKV-FMGC---58
Db 8 RGTLLPLLAALLQASVEASGEIALCKTGFPD-DVYSAVLPKDVHEGQPLLVKFSNCNRK 66

Qy 59 -----PQEPALFSTDDND--FTVRN-----GETVQER-----84
Db 67 RKVOYSESEFADFVKDSDGTVIYAFPLFAEQAKFLIYAQDKETQEKQVAVNLGREPT 126

Qy 85 ---RSLKERNPLK--IPFSK-----RILRRHRKRDVWVAPISVPENGKPPQRLNOLKSN 134
Db 127 LTEBPMKEPEHEIEIVPRLAKHSGALQKRDWIPPLNLPENSGRPPQELVIRSD 196

Qy 135 KDRTKLFYITGCAISPEGVFAVEKGTGWLKPLDREIAKYELPCHAVSNGAS 194
Db 187 RDKVLSRYSVTGEGADQPTGFIINPISGQSLSVTKPLDREIAFRAHAVIDNGNQ 246

Qy 195 VEDPMNISIIITDNDHKKPKFTQDTFRGSLVGLPCTSVNQVATDDEDAIYTVNGVA 254
Db 247 VENPIDIVINDNDNRPEFLHQVNGSVPEGKFTYVMTVTAIDADDP-NALNGMLR 305

Qy 255 YSHSQSPKPDHLMFTIHRSTGTISVISGLDREKVPETLTIQATMDGD---GSTTT 311
Db 306 YRILSQAPSTPSPNMFTINNETGDIITVAAGLDREKVVQYTLTIQATDMEGNPTYGLSNT 365

Qy 312 AVAVVELLDANDAPMDFDPKVEAHVPENAVHGVQRLVTVDLDAPNSPAWRATYLMGG 371
Db 366 ATAVITVDNDNDPPEPTAMTFYGEVEPENVIVANLVTIDKQPHTPANNAAYISGG 425

Qy 372 DDGDHETITTHPESNOGILITTRKGLDPEAKNOHTLYVEVTNEAPPVVKL---PTSTATIV 428
Db 426 DPTGRFAITDPSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKG:QHPPQSTATVS 485

Qy 429 VHVEDVNEAPVFPVPPSKVVEVEGPIPTGEPVVCVYTAEDPK-ENQKISYRIILRDPAGMLA 487
Db 486 VTVIDVNEPYPAPNPKIIRQEBGLHAGTMLITLTAQDPDRYMQQNIYTKLSDPANWLK 545

Qy 488 MDPDSQVTVAGTLDREDEQFVRNNIYVWVLAMDNQSPPTGTGTLLTLTLDVNDHGPV 547
Db 546 IDPVNGQITIAVLDRS-SPYVQNNIYNATFLASDNQIPPMGSGTGTLOLYLLDINDNAPQ 604

Qy 548 PEPRQITICNQSPVRHLNIT--DKDLSPTSPFOAQLTDDSDIY---WTAEVNE-EGDT 601
Db 605 VLPQEAETC-ETPEPNSINIAALDYIDNAGFFAFDELSPVTYKXWNT--INRLNGDF 661

Qy 602 VVLSLK-KFLKQDTYDVHLSLDHGN--KEQUTVIRATVCDCHG-----VETCPGPWKG 653
Db 662 AQLNLKIKLEAGIYEVPIIITDSGNPPKSNITSLRKVKVCQDSNGDCTDVRIVGAGLG 721

Qy 654 GFTLPVIGAVLALLFLLVLLLV-----SKRKKIKPELLLPEDDTDNDVYFGE 703
Db 722 -----TGAIILLCIIILLILVLMFVVMKRRKROKQKOLLIDPEDDVEDNLIKYE 775

Qy 704 EGGGEDQDYDITQHLRGLEAPV-----VLRNDVAPTIPTMYRPPA--NPDEIGN 756
Db 776 EGGGEDQDYLSQLQQDPTVEPDAIKPVGIRRLDERP-IHAEPQVYVRSAAHPGQDGD 834

Qy 757 FIENLKAANTDPTAPPYDTLTVDEYEGSGDAASLSLTSSASQDQDYDYLNGWSRF 816
Db 835 FINEGKAANDPTAPPYDLSLLVFDYEGSGTAGSLSSLNSSSGGQDYDYLNDWGPFR 894

Qy 817 KKLADMYGGGD 828
Db 895 KKLADMYGGGD 906

RESULT 15
IUCHCR
R-cadherin precursor - chicken
C/Species: Gallus gallus (chicken)
C/Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 16-Jun-2000
C/Accession: JH0424
R: Imazuka, H.; Miyatani, S.; Takeichi, M.
Neuron 7, 69-79, 1991
A/Title: R-cadherin: a novel Ca2+-dependent cell-cell adhesion molecule expressed in the
A/Reference number: JH0424; MUID: 91299341; PMID: 1712604
A/Accession: JH0424
A/Molecule type: mRNA
A/Residues: 1-913 <INU>
A/Cross-references: GB:D14459; GB:D00849; NID: g2222854; PIDN: BAA03356.1; PID: g2222855
A/Experimental source: retina
C/Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in cell-cell adhesion.
C/Superfamily: cadherin; cadherin repeat homology
C/Keywords: calcium binding; cell adhesion; duplication; glycoprotein; retina; transmembrane protein
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-166/Domain: propeptide #status predicted <PRO>
F:167-913/Product: R-cadherin #status predicted <MAT>
F:167-721/Domain: extracellular #status predicted <EXT>
F:169-274/Domain: cadherin repeat homology <CR1>
F:244-249/Region: cadherin binding #status predicted
F:277-389/Domain: cadherin repeat homology <CR2>
F:392-504/Domain: cadherin repeat homology <CR3>
F:507-612/Domain: cadherin repeat homology <CR4>
F:613-721/Domain: transmembrane #status predicted <TM>
F:722-753/Domain: transmembrane #status predicted <TM>
F:754-913/Domain: intracellular #status predicted <INT>
F:870-885/Region: serine-rich
F:280,409,554,629,658,699/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 38.8%; Score 1694; DB 1; Length 913;
Best Local Similarity 41.5%; Pred. No. 5e-100;

Search completed: September 21, 2004, 22:09:32
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:02:13 ; Search time 15 Seconds
(without alignments)
2877.744 Million cell updates/sec

Title: US-09-916-849A-1

Perfect score: 4369

Sequence: 1 MGLPRGPLASLLLLQVCWLQ.....NEWGRFRKLLADMYGGEDD 829

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	4369	100.0	829	CAD3_HUMAN	P22223 homo sapien
2	3550.5	81.3	822	CAD3_MOUSE	P10287 mus musculus
3	2573.5	58.9	732	CAD1_HUMAN	P33145 gallus gall
4	2420	55.4	882	CAD1_HUMAN	P12830 homo sapien
5	2388	54.7	884	CAD1_MOUSE	P09803 mus musculus
6	2388	54.7	886	CAD1_RAT	Q9r0t4 rattus norv
7	2375	54.3	880	CAD1_XENLA	P33148 xenopus lae
8	2371	54.3	884	CAD1_XENLA	P33152 xenopus lae
9	2323	53.2	491	CAD3_BOVIN	P19335 bos taurus
10	2274	52.0	887	CAD1_CHICK	P08841 gallus gall
11	2198	50.3	872	CAD1_XENLA	P30944 xenopus lae
12	1714	39.2	906	CAD2_HUMAN	P19022 homo sapien
13	1704.5	39.0	912	CAD2_CHICK	P10288 gallus gall
14	1697.5	38.9	906	CAD2_MOUSE	P15116 mus musculus
15	1694	38.8	913	CAD4_CHICK	P24503 gallus gall
16	1693	38.8	906	CAD2_RAT	Q92zy3 rattus norv
17	1691	38.7	877	CAD2_BOVIN	P19534 bos taurus
18	1681	38.5	893	CAD2_BRARE	Q90275 brachydanio
19	1655	37.9	913	CAD4_MOUSE	P39038 mus musculus
20	1621.5	37.1	916	CAD4_HUMAN	P55283 homo sapien
21	1616.5	37.0	906	CAD0_XENLA	P33147 xenopus lae
22	1609	36.8	905	CADN_XENLA	P20310 xenopus lae
23	1353.5	31.0	784	CAD6_MOUSE	P33146 mus musculus
24	1321.5	30.2	814	CAD7_HUMAN	P55291 homo sapien
25	1030.5	23.6	785	CAD7_HUMAN	Q9ulb5 homo sapien
26	1026	23.5	714	CADD_MOUSE	Q9wt5 mus musculus
27	1021.5	23.4	713	CADD_HUMAN	P55290 homo sapien
28	1014	23.2	902	DSC2_MOUSE	P55292 mus musculus
29	1007.5	23.1	785	CAD7_CHICK	Q90763 gallus gall
30	990	22.7	790	CAD6_CHICK	Q90762 gallus gall
31	983.5	22.5	792	CAD6_CHICK	O93319 gallus gall
32	982	22.5	790	CAD6_HUMAN	P55285 homo sapien
33	976.5	22.4	790	CADI_HUMAN	Q13634 homo sapien

34	976	22.3	712	1	CADD_CHICK	P33150 gallus gall
35	975	22.3	789	1	CAD9_HUMAN	Q9ulb4 homo sapien
36	974	22.3	789	1	CAD6_RAT	P55280 rattus norv
37	969.5	22.2	895	1	DSC3_MOUSE	P55850 mus musculus
38	966.5	22.1	796	1	CAD8_HUMAN	P55287 homo sapien
39	963	22.0	863	1	DSC2_BOVIN	P33245 bos taurus
40	958.5	21.9	790	1	CAD6_MOUSE	P97326 mus musculus
41	957	21.9	896	1	DSC3_HUMAN	Q14574 homo sapien
42	951.5	21.8	796	1	CAD8_MOUSE	P55288 mus musculus
43	950	21.7	788	1	CADA_HUMAN	Q9y6n8 homo sapien
44	948.5	21.7	901	1	DSC2_HUMAN	Q02487 homo sapien
45	940.5	21.5	896	1	DSC3_BOVIN	Q28060 bos taurus

ALIGNMENTS

RESULT 1	
CAD3_HUMAN	
ID CAD3_HUMAN	STANDARD; PRT; 829 AA.
AC P22223;	
DT 01-AUG-1991 (Rel. 19, Created)	
DT 01-AUG-1991 (Rel. 19, Last sequence update)	
DT 10-OCT-2003 (Rel. 42, Last annotation update)	
DE Cadherin-3 precursor (Placental-cadherin) (P-cadherin).	
GN CDH3 OR CDHP.	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX NCBI_TaxID:9606;	
RN [1]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=90009051; PubMed=2793940;	
RA Shimoyama Y., Yoshida T., Terada M., Shimosato Y., Abe O.,	
RA Hirotsashi S.;	
RT "Molecular cloning of a human Ca2+-dependent cell-cell adhesion	
RT molecule homologous to mouse placental cadherin: its low expression	
RT in human placental tissues.";	
RL J. Cell Biol. 109:1787-1794(1989).	
RN [2]	
RP SEQUENCE FROM N.A.	
RC TISSUE=Testis;	
RX MEDLINE=22388257; PubMed=12477932;	
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,	
RA Strausberg R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,	
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,	
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,	
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,	
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,	
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,	
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,	
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,	
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,	
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A.,	
RA Pahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Shevchenko Y.,	
RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,	
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,	
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,	
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,	
RA Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;	
RT "Generation and initial analysis of more than 15,000 full-length	
RT human and mouse cDNA sequences.";	
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).	
RN [3]	
RP DISEASE.	
RX MEDLINE=21470328; PubMed=11544476;	
RA Sprecher E., Bergman R., Richard G., Lurie R., Shalev S.,	
RA Petronius D., Shalata A., Ambinder Y., Leibur R., Perlman I., Cohen N.,	
RA Szargel R.;	
RT "Hypotrichosis with juvenile macular dystrophy is caused by a	
RT mutation in CDH3, encoding P-cadherin.";	
RL Nat. Genet. 29:134-136(2001).	
RN [4]	

RP VARIANT HJMD HIS-503.
 RX MEDLINE=22336734; PubMed=12445216;
 RA Indelman M., Bergman R., Lurie R., Richard G., Miller B.,
 RA Petronius D., Ciubutaro D., Leibu R., Sprecher E., causes
 RA "A missense mutation in CDH3, encoding p-cadherin, causes
 RT hypotrichosis with juvenile macular dystrophy.";
 RL J. Invest. Dermatol. 119:1210-1213(2002).
 CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
 CC They preferentially interact with themselves in a homophilic
 CC manner in connecting cells; cadherins may thus contribute to the
 CC sorting of heterogeneous cell types.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DISEASE: Defects in CDH3 are the cause of hypotrichosis with
 CC juvenile macular dystrophy (HJMD) [MIM:601553]. HJMD is a rare
 CC autosomal recessive disorder characterized by early hair loss
 CC heralding severe degenerative changes of the retinal macula and
 CC culminating in blindness during the second to third decade of
 CC life.
 CC -!- SIMILARITY: Contains 5 cadherin domains.
 CC -!- DATABASE: NAME=Mutations of the CDH3 gene;
 CC NOTE=Retina International's Scientific Newsletter;
 CC WWW="http://www.retina-international.com/sci-news/cdh3mut.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X63629; CAA45177.1; .
 DR EMBL; BC041846; AAH41846.1; .
 DR PIR; A33659; IJHUP.
 DR HSSP; P09803; 1SUH.
 DR Genew; HGNC:1762; CDH3.
 DR MIM; 114021; .
 DR MIM; 601553; .
 DR GO; GO:0007155; P:cell adhesion; TAS.
 DR InterPro; IPR002126; Cadherin.
 DR InterPro; IPR000233; Cadherin_C_term.
 DR Pfam; PF00028; cadherin; 5.
 DR Pfam; PF01049; Cadherin_C term; 1.
 DR PRINTS; PR00205; CADHERIN.
 DR SMART; SM00112; CA; 4.
 DR PROSITE; PS00232; CADHERIN_1; 3.
 DR PROSITE; PS0268; CADHERIN_2; 5.
 DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
 KW Signal; Disease mutation; Vision.
 FT SIGNAL 1 24
 FT PROPEP 25 107
 FT CHAIN 108 829
 FT DOMAIN 108 829 CADHERIN-3.
 FT TRANSMEM 108 854 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 655 677 POTENTIAL.
 FT CYTOPLASMIC 678 829 POTENTIAL).
 FT CADHERIN 1. 108 215
 FT CADHERIN 2. 216 328
 FT CADHERIN 3. 329 440
 FT CADHERIN 4. 441 546
 FT CADHERIN 5. 547 800
 FT SER-RICH. 785 800
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 566 566 R -> H (in HJMD).
 FT VARIANT 503 503 /FTD=VAR_015422.
 FT CONFLICT 237 237 V -> M (IN REF. 2).
 FT CONFLICT 563 563 H -> Q (IN REF. 2).
 FT CONFLICT 829 829 AA; 91427 MW; E503CFE5D981F1 CRC64;
 SQ SEQUENCE 829 AA; 91427 MW; E503CFE5D981F1 CRC64;
 Query Match 100.0%; Score 4369; DB 1; Length 829;
 Best Local Similarity 100.0%; Pred. No. 2.7e-264;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RX MEDLINE=88111554; PubMed=3428270;

QY 1 MGLPRGPLASLLQLQVWLOCAASEPCRAVPREAEVTLERAGABQERGOALGKVFMCSPG 60
 DB 1 MGLPRGPLASLLQLQVWLOCAASEPCRAVPREAEVTLERAGABQERGOALGKVFMCSPG 60
 QY 61 CEPALFSTDNDDFTVRNGETQERSLKERNLPIFKSKILRRHKRDWVAPISVPENG 120
 DB 61 CEPALFSTDNDDFTVRNGETQERSLKERNLPIFKSKILRRHKRDWVAPISVPENG 120
 QY 121 KGPFFQRLNQLSKNDRDKTKIFYITGPGADSPPEGVFAVEKETGWLILNKLPLDREIAK 180
 DB 121 KGPFFQRLNQLSKNDRDKTKIFYITGPGADSPPEGVFAVEKETGWLILNKLPLDREIAK 180
 QY 181 YELFGHAYSENGASVEDPMNISIIITDQNDHKPFTQDTRGSLGVLPGTSVMQVTAT 240
 DB 181 YELFGHAYSENGASVEDPMNISIIITDQNDHKPFTQDTRGSLGVLPGTSVMQVTAT 240
 QY 241 DEDDAIYTYNGVWVAYS IHSQEPKDPHDLMTIHRSTGTISVISGLDREKVPYTLTQA 300
 DB 241 DEDDAIYTYNGVWVAYS IHSQEPKDPHDLMTIHRSTGTISVISGLDREKVPYTLTQA 300
 QY 301 TDMGDGSGTTTAVAVVEILDANDNAPMDPKQYEAHVPENAVGHVQRLTVDLDAENSP 360
 DB 301 TDMGDGSGTTTAVAVVEILDANDNAPMDPKQYEAHVPENAVGHVQRLTVDLDAENSP 360
 QY 361 AWRATYLLMGDDGDHFTITTHPESNOGILTRKGLDFEAKNQHTLYVEVTNEAPFVLKL 420
 DB 361 AWRATYLLMGDDGDHFTITTHPESNOGILTRKGLDFEAKNQHTLYVEVTNEAPFVLKL 420
 QY 421 PTSTATIIVHVEDVNEAPVFPVPPSKVVEVQGIPTGEPVCYVTAEDDPKKNOKISYILR 480
 DB 421 PTSTATIIVHVEDVNEAPVFPVPPSKVVEVQGIPTGEPVCYVTAEDDPKKNOKISYILR 480
 QY 481 DPAGWLAMPDPSGGQVAVTGLDREDEQVRNNIYEVWVLANMGNSPPTTGTLTLLTLD 540
 DB 481 DPAGWLAMPDPSGGQVAVTGLDREDEQVRNNIYEVWVLANMGNSPPTTGTLTLLTLD 540
 QY 541 VNDHGFVPEPRQITTCNQSPVRHVNLITDKLSPTSPFQAQLTDDSDIYWTAEVNEGD 600
 DB 541 VNDHGFVPEPRQITTCNQSPVRHVNLITDKLSPTSPFQAQLTDDSDIYWTAEVNEGD 600
 QY 601 TVVLSIKKFLKQDITYVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGFIPLVL 660
 DB 601 TVVLSIKKFLKQDITYVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGFIPLVL 660
 QY 661 GAVLALLFLLVLLVLLVLRKKIKLEPLLPEDDTRDNVFFYGGEGGEDQDITQLHR 720
 DB 661 GAVLALLFLLVLLVLLVLRKKIKLEPLLPEDDTRDNVFFYGGEGGEDQDITQLHR 720
 QY 721 GLEARPEVVLNDVAPTIIPTPMYRPRANDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 DB 721 GLEARPEVVLNDVAPTIIPTPMYRPRANDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 QY 781 DYEGSGSDAASLSLTSASDQDQDYLYNEWGSRFKKLADMYGGEDD 829
 DB 781 DYEGSGSDAASLSLTSASDQDQDYLYNEWGSRFKKLADMYGGEDD 829

RESULT 2
 CAD3_MOUSE
 ID_CAD3_MOUSE STANDARD; PRT; 822 AA.
 AC P10387; Q61465;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cadherin-3 precursor (Placental-cadherin) (p-cadherin).
 GN CDH3 OR CDHP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88111554; PubMed=3428270;

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE B-cadherin precursor (K-CAM protein) (Fragment).
 GN K-CAM.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=92107987; PubMed=1763068;
 RX Sorokin B.C., Gallin W.J., Edelman G.M., Cunningham B.A.;
 RA "Genes for two calcium-dependent cell adhesion molecules have similar
 RT structures and are arranged in tandem in the chicken genome.";
 RT Proc. Natl. Acad. Sci. U.S.A. 88:11545-11549(1991).
 RL [2]
 RN SEQUENCE OF 7-732 FROM N.A.
 RP TISSUE=Embryonic brain;
 RX MEDLINE=91225083; PubMed=2026653;
 RA Napolitano E.W., Venstrom K., Wheeler E.F., Reichardt L.F.;
 RT "Molecular cloning and characterization of B-cadherin, a novel chick
 RT cadherin.";
 RL J. Cell Biol. 113:893-905(1991).
 CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
 CC They preferentially interact with themselves in a homophilic
 CC manner in connecting cells; cadherins may thus contribute to the
 CC sorting of heterogeneous cell types. B-cadherin may have important
 CC functions in neurogenesis, in at least some epithelia, and in
 CC embryogenesis.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: Expressed in a wide variety of tissues.
 CC -!- SIMILARITY: Contains 5 cadherin domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M81894; AAA4829.1; -;
 DR EMBL; X58518; CA41408.1; -;
 DR PIR; A41634; IJCHCB.
 DR HSSP; P09803; 1SUH.
 DR InterPro; IPR002126; Cadherin.
 DR InterPro; IPR002233; Cadherin_C_term.
 DR Pfam; PF00028; cadherin; 5;
 DR Pfam; PF01049; Cadherin_C_term; 1.
 DR PRINTS; PR00205; CADHERIN.
 DR PROSITE; PS00232; CADHERIN_1; 3.
 DR PROSITE; PS0268; CADHERIN_2; 5.
 DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 FT NON TER 1 1
 FT PROTEP 1 6
 FT CHAIN 7 732
 FT DOMAIN 6 554
 FT TRANSMEM 555 580
 FT DOMAIN 581 732
 FT CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 1. 114
 FT CADHERIN 1. 115
 FT DOMAIN 2. 227
 FT CADHERIN 2. 228
 FT DOMAIN 3. 339
 FT CADHERIN 3. 340
 FT DOMAIN 4. 443
 FT CADHERIN 4. 444
 FT DOMAIN 5. 554
 FT CADHERIN 5. 554
 FT SER-RICH. 689 702
 FT DOMAIN 137 137
 FT CARBOHYD N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 410 410
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT M -> V (IN REF. 2).
 FT CONFLICT 414 414
 FT SEQUENCE 732 AA; 80613 MW; 091D59A6A16CBD45 CRC64;

Query Match 58.9%; Score 2573.5; DB 1; Length 732;
 Best Local Similarity 66.2%; Fred. No. 1.3e-152;
 Matches 485; Conservative 103; Mismatches 138; Indels 7; Gaps 5;

QY 102 LRRHKDWWVAPISVPEKGKPPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVE 161
 DB 1 LRRQRDWWIPPIKVPENERGPPFANLVQIAKSNDRRAKIFYSITGGADAPPGIETIE 60
 QY 162 KETGMLLNKLPDREIEIAKELFCHAVSENGASVEDPMN:SIIVTDQNDHKPKFTQTFR 221
 DB 61 KETGMKVTQPLDREHINKYHLYSHAVSENGKPEEPMEIIVTVDQNDKNKQPQTQSVFR 120
 QY 222 GSVLEGVLPCTSVNQVATDDEDAIYNGVAVYSIHSQEPKDPHDLMEFTHRSTGLISV 281
 DB 121 GSVPEGALPGTSMVRVNAIDADDVETNGVIAYSIISQEPREPHMFTVNRATGLSV 180
 QY 282 ISSGLDRKVEPYTLTIQATDMGDGSGTTTAVAVVEILDANDNAPMFDPOKYEAHVPENA 341
 DB 181 IASGLDRERVREYTLTMOAADLGGGLTTTALAIEITVDNDNAPEFDPKTYEAAVENE 240
 QY 342 VGHEVQRLTVTDLDAIPNPAWRATYILMGDDGDGHFTITTHPESNOGILLTRKGLDEAK 401
 DB 241 AELEVARLATTLDEPHTPAWRAVYSIVRGEGGAFTITDPASNEGVLRITAKGLDEAK 300
 QY 402 NCHTLVYVETNEAPFVLKLTSTATIVVHVEDVNEAPVFVPPSKVVEVQEGIPTGEPCV 461
 DB 301 RQFVLAHVAVNEAPFAIKLPTATATVMVSDVNEAPVDFPPLAQVPEDVPLGQPLAS 360
 QY 462 YTAEDPK-ENQKISYRILRDPAGWLAMPDSGQVAVGTLDREDEQFVRNIIYEVWLA 520
 DB 361 YTAQDPDRAQQRIKYVMGSDPAGWLAVHPENGLITAREQLDRE-SPFTKNSTYNAVLLA 419
 QY 521 MDNGSPPTGTGTLTLTLDVNDHGPVPEPQITICNQSPVRHVLNITDKDLSPTGPFQ 580
 DB 420 VDDGPPATGTGTLTLTLDVNDHGPVPEPQITICNQSPVPQVLTITDRDLPNTGPF 479
 QY 581 AQLTDDSDIYNTAEVNEEGDTVLSLKKFLKODTYDVHLSLSDHGNKEQLTVIRATVCDC 640
 DB 480 AELSHGSGDSWAVEVNGGDTVALWLTPELQNLVSVVRLFDQKQDQVTVIRAQVDC 539
 QY 641 HGVETC---PGPWKG-CFILPVIGAVLALFLALLVLLVLRKKRKEPELLEDTRD 696
 DB 540 QQRVESCAQKPRVDTGVFVIAVLGAVLALLVLLVLRKKRKEPELLEDTRD 599
 QY 697 NVFYEGEGGGEEDQYDITQLHRLGLARPVLENDVAPTIPTPMYRPPANPDEIGN 756
 DB 600 NIFYEGEGGGEEDQYDLSQLHRLGLARPEVI-RNDVAPPLMAAPQYRPPANPDEIGN 658
 QY 757 FIENLKAANTDPTAPPYDITLLVFDYEGSGDAASLSLTSASDODDYLNWGSRF 816
 DB 659 FIDENLKAADTPTAPPYDLSLLVFDYEGGSEATSLSSLSSASDQDQDYLNWGNRF 718
 QY 817 KKLADMYGGGEDD 829
 DB 719 KKLAELYGGGEDE 731

RESULT 4

CAD1_HUMAN STANDARD; PRT; 882 AA.
 ID_CAD1_HUMAN Q13799; Q14216; Q15855; Q16194;
 AC P12670; Q13799; Q14216; Q15855; Q16194;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Epithelial-cadherin precursor (E-cadherin) (Uvomorulin) (cadherin-1)
 DE (CAM 120/80).
 GN CDH1 OR UVO OR CDHE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=93211394; PubMed=8459805;
 RA Bussemakers M.J.G., Mees S.G.M., van Bokhoven A., Debruyne F.M.J.,

RA Schalken J.A.;
RT "Molecular cloning and characterization of the human E-cadherin
RL cDNA.";
RN Mol. Biol. Rep. 17:123-128(1993).
[2]
RP SEQUENCE FROM N.A.
RA Kelker W., Warda A., Oda T., Hirohashi S., Kemler R., Birchmeier W.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94242050; PubMed=8185635;
RA Rimm D.L., Morrow J.S.;
RT "Molecular cloning of human E-cadherin suggests a novel subdivision of
RL the cadherin superfamily.";
RN Biochem. Biophys. Res. Commun. 200:1754-1761(1994).
[4]
RP SEQUENCE OF 56-882 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=95324920; PubMed=7601454;
RA Bex G., Staes K., van Hengel J., Molenans F., Bussemakers M.J.G.,
RA van Bokhoven A., van Roy F.;
RT "Cloning and characterization of the human invasion suppressor gene
RL E-cadherin (CDH1).";
RN Genomics 26:281-289(1995).
[5]
RP SEQUENCE OF 172-311 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=89031725; PubMed=3263290;
RA Mansouri A., Spurr N., Goodfellow P.N., Kemler R.;
RT "Characterization and chromosomal localization of the gene encoding
RL the human cell adhesion molecule uvomorulin.";
RN Differentiation 38:67-71(1988).
[6]
RP SEQUENCE OF 265-392 FROM N.A.
RC TISSUE=Liver;
RA Frixen U.H.;
RL Submitted (MAR-1990) to the EMBL/GenBank/DBJ databases.
[7]
RP SEQUENCE OF 1-16 FROM N.A.
RX MEDLINE=94380041; PubMed=8093045;
RA Bussemakers M.J.G., Girolidi L.A., van Bokhoven A., Schalken J.A.;
RT "Transcriptional regulation of the human E-cadherin gene in human
RL prostate cancer cell lines: characterization of the human E-cadherin
RN gene promoter.";
RL Biochem. Biophys. Res. Commun. 203:1284-1290(1994).
[8]
RP SEQUENCE OF 1-16 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=95365379; PubMed=7543680;
RA Yoshiura K., Kanai Y., Ochiai A., Shimoyama Y., Sugimura T.,
RA Hirohashi S.;
RT "Silencing of the E-cadherin invasion-suppressor gene by CpG
RL methylation in human carcinomas.";
RN Proc. Natl. Acad. Sci. U.S.A. 92:7416-7419(1995).
[9]
RP INTERCHAIN DISULFIDE BOND.
RX MEDLINE=21975237; PubMed=11856755;
RA Makagiansar I.T., Nguyen P.D., Ikesue A., Kuczer A., Dentler W.,
RA Urbauer J.L., Galeva N., Alterman M., Sahaan T.J.;
RT "Disulfide bond formation promotes the cis- and trans-dimerization of
RL the E-cadherin-derived first repeat.";
RN J. Biol. Chem. 277:16002-16010(2002).
[10]
RP DISEASE.
RX MEDLINE=99406500; PubMed=10477433;
RA Guilford P.J., Hopkins J.B.W., Grady W.M., Markowitz S.D., Willis J.,
RA Lynch H., Rajput A., Wiesner G.L., Lindor N.M., Burgart L.J.,
RA Toro T.R., Lee D., Limacher J.-M., Shaw D.W., Findlay M.P.N.,
RA Reeve A.E.;
RT "E-cadherin germline mutations define an inherited cancer syndrome
RL dominated by diffuse gastric cancer.";
RN Hum. Mutat. 14:249-255(1999).
[11]
RP REVIEW ON VARIANTS.
RX MEDLINE=98415721; PubMed=9744472;
RA Bex G., Becker K.-F., Hoefler H., van Roy F.;
RT "Mutations of the human E-cadherin (CDH1) gene.";
RN Hum. Mutat. 12:226-237(1998).
[12]
RP SEQUENCE OF 337-476 FROM N.A., AND VARIANTS ALA-370 AND ASP-473.
RX MEDLINE=94306394; PubMed=8033105;
RA Becker K.-F., Atkinson M.J., Reich U., Becker I., Nekarda H.,
RA Siewert J.R., Hoefler H.;
RT "E-cadherin gene mutations provide clues to diffuse type gastric
RL carcinomas.";
RN Cancer Res. 54:3845-3852(1994).
[13]
RP VARIANT LOBULAR BREAST CARCINOMA SER-315.
RX MEDLINE=95049851; PubMed=7961105;
RA Kanai Y., Oda T., Tsuda H., Ochiai A., Hirohashi S.;
RT "Point mutation of the E-cadherin gene in invasive lobular carcinoma
RL of the breast.";
RN Jpn. J. Cancer Res. 85:1035-1039(1994).
[14]
RP VARIANTS GYNECOLOGIC CANCERS THR-617; VAL-711 AND GLY-838.
RX MEDLINE=94355985; PubMed=8075649;
RA Risinger J.I., Berchuck A., Kohler M.F., Boyd J.;
RT "Mutations of the E-cadherin gene in human gynecologic cancers.";
RN Nat. Genet. 7:98-102(1994).
[15]
RP VARIANT GASTRIC ADENOCARCINOMA 274-GLY--PRO-277 DEL.
RX MEDLINE=94173928; PubMed=8127895;
RA Oda T., Kanai Y., Oyama T., Yoshiura K., Shimoyama Y., Birchmeier W.,
RA Sugimura T., Hirohashi S.;
RT "E-cadherin gene mutations in human gastric carcinoma cell lines.";
RN Proc. Natl. Acad. Sci. U.S.A. 91:1858-1862(1994).
[16]
RP VARIANT GASTRIC CARCINOMA PRO-193.
RX MEDLINE=96390918; PubMed=8797891;
RA Muta H., Noguchi M., Kanai Y., Ochiai A., Nawata H., Hirohashi S.;
RT "E-cadherin gene mutations in signet ring cell carcinoma of the
RN stomach.";
RL Jpn. J. Cancer Res. 87:843-848(1996).
[17]
RP VARIANTS GASTRIC CARCINOMA ASP-400 DEL AND 418-ASP--PHE-423 DEL.
RX MEDLINE=97197648; PubMed=9045944;
RA Tamura G., Sakata K., Nishizuka S., Maesawa C., Suzuki Y., Iwaya T.,
RA Terashima M., Saito K., Satodate R.;
RT "Inactivation of the E-cadherin gene in primary gastric carcinomas and
RL gastric carcinoma cell lines.";
RN Jpn. J. Cancer Res. 87:1153-1159(1996).
[18]
RP VARIANT THYROID TUMOR THR-592.
RX MEDLINE=97138061; PubMed=8985087;
RA Soares P., Bex G., van Roy F., Sobrinho-Simoes M.;
RT "E-cadherin gene alterations are rare events in thyroid tumors.";
RN Int. J. Cancer 70:32-38(1997).
[19]
RP VARIANTS ASP-336 AND ILE-470.
RX MEDLINE=98196671; PubMed=9537325;
RA Guilford P.J., Hopkins J.B.W., Harraway J., McLeod M., McLeod N.,
RA Harawita P., Taite H., Scullar R., Miller A., Reeve A.E.;
RT "E-cadherin germline mutations in familial gastric cancer.";
RN Nature 392:402-405(1998).
[20]
RP VARIANTS HDGC GLY-244 AND ALA-487.
RX MEDLINE=99253140; PubMed=10319582;
RA Yoon K.-A., Ku J.-L., Yang H.-K., Kim W.H., Park S.Y., Park J.-G.;
RT "Germline mutations of E-cadherin gene in Korean familial gastric
RL cancer patients.";
RN J. Hum. Genet. 44:177-180(1999).
[21]
RP VARIANT COLORECTAL CANCER ALA-340.
RX MEDLINE=20357134; PubMed=10896919;
RA Kim H.C., Wheeler J.M.D., Kim J.C., Ilyas M., Beck N.E., Kim B.S.,

RA Park K.C., Bodmer W.F.;
RT "The E-cadherin gene (CDH1) variants T340A and L599V in gastric and
RL colorectal cancer patients in Korea.";
RN Gut 47:262-267(2000).
RP
RX VARIANT ALA-270.
RX MEDLINE=21562463; PubMed=11705864;
RA Ikonen T., Matikainen M., Mononen N., Hyyinen E.R., Helin H.J.,
RA Tomola S., Tammela T.L., Pukkala E., Schleutker J., Kallioniemi O.P.,
RA Kivisto P.A.;
RT "Association of E-cadherin germ-line alterations with prostate
RT cancer.";
RL Clin. Cancer Res. 7:3465-3471(2001).
RN
RN VARIANT THR-592.
RX MEDLINE=21446527; PubMed=11562785;
RA Salahshor S., Hou H., Diep C.B., Loukola A., Zhang H., Liu T.,
RA Chen J., Iselius L., Rubio C., Lothe R.A., Aaltonen L., Sun X.F.,
RA Lindmark G., Lindblom A.;
RT "A germline E-cadherin mutation in a family with gastric and colon
RT cancer.";
RL Int. J. Mol. Med. 8:439-443(2001).
RN
RN VARIANT HDGC ALA-340.
RX MEDLINE=22021361; PubMed=11968083;
RA Oliveira C., Bordin M.C., Grehen N., Huntsman D., Suriano G.,
RA Machado J.C., Kiviluoto T., Aaltonen L., Jackson C.E., Seruca R.,
RA Caldas C.;
RT "Screening E-cadherin in gastric cancer families reveals germline
RT mutations only in hereditary diffuse gastric cancer kindred.";
RL Hum. Mutat. 19:510-517(2002).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types. E-cadherin has a potent
CC invasive suppressor role. It is also a ligand for integrin alpha-
CC E/beta-7.

Query Match 55.4%; Score 2420; DB 1; Length 882;
Best Local Similarity 55.1%; Pred. No. 5.8e-143;
Matches 483; Conservative 112; Mismatches 222; Indels 60; Gaps 13;

QY 9 ASLLQLQV-CWLOCASEPCEAFREAEVLEAGAEQEPQALGKV-FMGCPG-QEPAL 65
DB 10 ALLLLQVSSL-CQEPCHPGDAESYFTVPRHLERGRVLFGRVNFEDCTGRTAY 68
QY 66 PSTNDPFTVRNGTIVERRSLKERNP----- 92
DB 69 FSLDT-RFKVGTGVIITKRLPRHNPQIHLVYVANDSTVRKFTSKVTLNVTGHHRRPP 127
QY 93 -----LKIFP-SKILRHKEDVWVAPISVPEKGGPPQRLNQLKNDKRDTK 140
DB 128 HQASVSGIQAEALLTFPNSPCLRQKQDWIPPIPCENEGKPPKLVQIKSKDKEG 167
QY 141 IFYSITGEGADSPGEPFAVEKETGMLLNKPLDREBIAYELFPHAVSNGASVEDPMN 200
DB 188 VFYSITGQADTPVGVFIHERETGMLKTEPLDRERIATYTLFSAVSSNGNAVEDPM 247
QY 201 ISIVTQNDHKPKFTDTRGSLVLEGLPQTSVQVATDDEDAIYTVNGVAVYSHQ 260
DB 248 ILIIVTQNDKPKFTQEVFGKSVMEALPQTSVMEVATDADDVNTYNAATYILSQ 307
QY 261 EPKDPHLMFTIHRSTGTISVSSGLDREKVPYETLTIOATMDGSGSTTAVAVVILD 320
DB 308 DPFLPDKNMFTINRTGIVSVVTTGLDRSFSTYTLVQVADLQGEGLSTATAVITVD 367
QY 321 ANDNAPMFDPOKFAHPENAVGHEVQLTVTDLDPNSPAWRATYILMGDDGDHFTT 380
DB 368 TNDNPPFTYKGVFENAVNVIITLKVTDADNTPAWAVTIL-NDGGQGVVT 426
QY 381 THPESNGILTRKGLDFEAKNQHTLVYVNTNEAPVFLKLPSTATVHVHEDVNEAPVF 440
DB 427 TNPVNDGILKTAKGLDFEAKQYILHVAIVNVVFPFVSLTITATVTVDLVDVNEAPIF 486

QY 441 VPPSKVVEVQSGIPTGEPVVCVYTAEDPK-ENQKISYRILRDPAGWLAMDPSQGVTAVG 499
DB 487 VPPEKRVESSEDFGVQGEITSYTAQEPDTEMEQKITVIRWDRDTANWLEINPDTGAISTRA 546
QY 500 TLDREDEQFVNNIYVWVLAMDNGSPPTTGTGLTLLTLDVNDHGPVPEPRTICNQOS 559
DB 547 ELDRDEDFHVNKNSTYATLIDTNGSPVATGTGILLILSDVNDNAPIPRTIFFCERN 606
QY 560 PYRHVINITKDLSPHTSPFOAQLTDSDIYWTAEVNE-EGDTVVLSLKFLKQDTYDVH 618
DB 607 PKPQVINIIDADLPENTSPFTAEILTHGASANNTIYNDPTQESILKPKQWALEVGYKIN 666
QY 619 LLSLDHGNKEQLTVIRATVCDCHGVETC--PGWKGKF-----LPLVGLAVLLELLLV 672
DB 667 LKLMNQNKQDQVTTLEVSVCCEGAAGVCRKAQVEAGLQIPAILGILGLALLILL 726
QY 673 LLLLVRKXKIKKEPLLPPEDDTRDNVYFYEGEGGEDQDYDITQLHRGLEARPEVVLN 732
DB 727 LLLFLRRRAVVKPELLPPEDDTRDNVYFYEGEGGEDQDFDLSQLHRGLDARPEVT-RN 785
QY 733 DVAPTIIPTPMYRPRPNDPEIGNFIENLKAANTDTPAPPYDILLVDFYEGSGSDASL 792
DB 786 DVAPTIIPTPMYRPRPNDPEIGNFIENLKAANTDTPAPPYDILLVDFYEGSGSDASL 845
QY 793 SSLTSSASDQDQDYDLNMGSRFKKLADMYGGGEDD 829
DB 846 SSLNSESSEDKQDQDYDLNMGSRFKKLADMYGGGEDD 882

RESULT 5
CAD1 MOUSE STANDARD; PRT; 884 AA.
ID CAD1 MOUSE
AC P09603; Q61377;
DT 01-VAR-1989 (Rel. 10, Created)
DT 01-VAR-1989 (Rel. 10, last sequence update)
DT 15-VAR-2004 (Rel. 43, last annotation update)
DE Epithelial-cadherin precursor (E-cadherin) (Uvomorulin) (Cadherin-1)
DE (ARC-1).
GN CDH1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR;
RX MEDLINE=87315445; PubMed=3498123;
RX Nagafuchi A., Shirayoshi Y., Okazaki K., Yasuda K., Takeichi M.;
RT "Transformation of cell adhesion properties by exogenously introduced
RT E-cadherin cDNA.";
RL Nature 329:341-343(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/Sv;
RX MEDLINE=92093614; PubMed=1754391;
RX Ringwald M., Baribault H., Schmidt C., Kemler R.;
RT "The structure of the gene coding for the mouse cell adhesion
RT molecule uvomorulin.";
RL Nucleic Acids Res. 19:6533-6539(1991).
RN [3]
RP SEQUENCE OF 174-884 FROM N.A., AND SEQUENCE OF 157-181.
RX MEDLINE=88111553; PubMed=3501370;
RX Ringwald M., Schun R., Vestweber D., Eistetter H., Lottspeich F.,
RX Engel J., Doelz R., Jaehning F., Epplen J., Mayer S., Mueller C.,
RX Kemler R.;
RT "The structure of cell adhesion molecule uvomorulin. Insights into
RT the molecular mechanism of Ca2+-dependent cell adhesion.";
RL EMBO J. 6:3647-3653(1987).
RN [4]
RP SEQUENCE OF 1-15 FROM N.A.
RX MEDLINE=92107977; PubMed=1763063;
RX Behrens J., Loewrick O., Klein-Hitpass L., Birchmeier W.;

RT "The E-cadherin promoter: functional analysis of a G.C-rich region
 RT and an epithelial cell-specific palindromic regulatory element.",
 RL Proc. Natl. Acad. Sci. U.S.A. 86:11495-11499(1991).
 RN [5]
 RP DEVELOPMENTAL STAGE.
 RP STRAIN=C57BL/6; TISSUE=Testis;
 RC MEDLINE=97033837; PubMed=8879495;
 RX Munro S.B., Blaschuk O.W.;
 RA "A comprehensive survey of the cadherins expressed in the testes of
 RT fetal, immature, and adult mice utilizing the polymerase chain
 RT reaction.",
 RL Biol. Reprod. 55:822-827(1996).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 157-370.
 RP MEDLINE=96176249; PubMed=8598933;
 RX Nagar B., Overduin M., Ikura M., Rini J.M.;
 RA "Structural basis of calcium-induced E-cadherin rigidification and
 RT dimerization.",
 RL Nature 380:360-364(1996).
 RN [7]
 RP STRUCTURE BY NMR OF 157-260.
 RP MEDLINE=96271285; PubMed=8785495;
 RX Overduin M., Tong X.J., Kay C.M., Ikura M.;
 RA "1H, 15N and 13C resonance assignments and monomeric structure of the
 RT amino-terminal extracellular domain of epithelial cadherin.",
 RL J. Biomol. NMR 7:173-189(1996).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS) OF 577-728 IN COMPLEX WITH
 RP CTNNB1, AND PHOSPHORYLATION.
 RX MEDLINE=21246507; PubMed=11348595;
 RA Huber A.H., Weis W.I.;
 RT "The structure of the beta-catenin/E-cadherin complex and the
 RT molecular basis of diverse ligand recognition by beta-catenin.",
 RL Cell 105:391-402(2001).
 CC -1- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
 CC They preferentially interact with themselves in a homophilic
 CC manner in connecting cells; cadherins may thus contribute to the
 CC sorting of heterogeneous cell types. E-cadherin is a ligand for
 CC integrin alpha-E/beta-7.
 CC -1- SUBUNIT: Homodimer. Binds CTNNB1.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: Non-neutral epithelial tissues.
 CC -1- DEVELOPMENTAL STAGE: In the testis, expression is highest in fetal
 CC gonad, then decreases 5-fold in newborn. Detectable in 7-day-old
 CC but not in 21-day-old or adult.
 CC -1- SIMILARITY: Contains 5 cadherin domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X06115; CAA29488.1; -
 DR EMBL; X60961; CAA43292.1; -
 DR EMBL; X60962; CAA43292.1; JOINED.
 DR EMBL; X60963; CAA43292.1; JOINED.
 DR EMBL; X60964; CAA43292.1; JOINED.
 DR EMBL; X60965; CAA43292.1; JOINED.
 DR EMBL; X60966; CAA43292.1; JOINED.
 DR EMBL; X60967; CAA43292.1; JOINED.
 DR EMBL; X60968; CAA43292.1; JOINED.
 DR EMBL; X60969; CAA43292.1; JOINED.
 DR EMBL; X60970; CAA43292.1; JOINED.
 DR EMBL; X60971; CAA43292.1; JOINED.
 DR EMBL; X60972; CAA43292.1; JOINED.
 DR EMBL; X60973; CAA43292.1; JOINED.
 DR EMBL; X60974; CAA43292.1; JOINED.
 DR EMBL; X60975; CAA43292.1; JOINED.
 DR EMBL; X08339; CAA29645.1; -
 DR EMBL; M81449; AAA37352.1; -

DR PIR; S04528; IJMSCE.
 DR PIR; S34438; S34438.
 DR PDB; 1EDH; 11-JAN-97.
 DR PDB; 1SUH; 11-JUL-96.
 DR PDB; 117W; 09-MAY-01.
 DR PDB; 117X; 16-MAY-01.
 DR PDB; 1PF5; 23-AUG-00.
 DR MGD; MGI:89354; Cdh1.
 DR GO; GO:0005886; C:Plasma membrane; IDA.
 DR GO; GO:0005503; F:calcium ion binding; IDA.
 DR GO; GO:0042804; F:protein homooligomerization activity; IDA.
 DR InterPro; IPR002126; Cadherin.
 DR InterPro; IPR00233; Cadherin_C term.
 DR Pfam; PF01049; Cadherin_C term; 1.
 DR PRINTS; PR00205; CADHERIN.
 DR SMART; SM00112; CA; 4.
 DR PROSITE; PS00232; CADHERIN_1; 3.
 DR PROSITE; PS0268; CADHERIN_2; 4.
 DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
 KW Signal; Phosphorylation; 3D-structure.
 FT SIGNAL 1 23
 FT PROPEP 24 156
 FT CHAIN 157 884
 FT DOMAIN 157 709
 FT TRANSMEM 710 733
 FT DOMAIN 734 884
 FT DOMAIN 157 264
 FT DOMAIN 265 377
 FT DOMAIN 378 488
 FT DOMAIN 489 595
 FT DOMAIN 596 699
 FT DOMAIN 840 855
 FT MOD_RES 840 840
 FT MOD_RES 842 842
 FT MOD_RES 848 848
 FT CARBOHYD 560 560
 FT CARBOHYD 639 639
 FT CONFLICT 287 267
 FT CONFLICT 272 272
 FT STRAND 163 166
 FT TURN 167 168
 FT STRAND 175 179
 FT HELIX 183 186
 FT TURN 187 187
 FT STRAND 190 195
 FT STRAND 197 197
 FT TURN 198 200
 FT STRAND 201 201
 FT TURN 204 205
 FT STRAND 207 209
 FT TURN 211 213
 FT STRAND 215 218
 FT TURN 224 226
 FT STRAND 229 238
 FT TURN 239 240
 FT STRAND 243 243
 FT STRAND 248 255
 FT STRAND 263 264
 FT STRAND 268 274
 FT TURN 275 276
 FT TURN 279 280
 FT STRAND 282 285
 FT STRAND 288 289
 FT TURN 294 296
 FT TURN 299 300
 FT STRAND 303 310
 FT STRAND 319 321
 FT TURN 323 325
 FT STRAND 327 330
 FT TURN 337 339
 FT STRAND 342 350
 FT TURN 351 351

FT	MOD_RES	844	844	PHOSPHORYLATION (BY SIMILARITY).	
FT	MOD_RES	850	850	PHOSPHORYLATION (BY SIMILARITY).	
FT	CARBOHYD	562	562	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	641	641	N-LINKED (GLCNAC. . .) (POTENTIAL).	
SQ	SEQUENCE	886 AA;	98714 MM;	A9AEE28EB797A547 CRC64;	
Query Match		54.7%;	Score 2388;	DB 1;	Length 886;
Best Local Similarity		53.8%;	Pred. No. 5.8e-141;		
Matches 477;	Conservative 122;	Mismatches 212;	Indels 76;	Gaps 14;	
QY	9	ASLLLLQV-CWLQC----	AASEPCRAVFREAEVTLLEAGGAEQEPGQALGKV-FMGCPGOE	62	
DB	10	ALLLLIQLVSSWL-CQOPESESDSCRPGFSSEVYTLVPERHLERGHILGRVKEGCTGRP	68		
QY	63	PALFSTDNDDFTVRNGETVQERSLK-----	88		
DB	69	RTAFSEDSRFKVTGDVITVTRHKLHKLHKLKLETSFLVHAWDSYRKLSKVTLSLGHGHH	128		
QY	89	-----ERNP-LKIPPS-KRIILRRKRDWVAPISVPENGGKPPQRLNOLKSNKDR	137		
DB	129	RHHHRDPVESNELLTTPSFHQGLRRQKQDWVIPPINCENQKGEPPQRLVQIKSNRDK	198		
QY	138	DTKIFVITGPGADGPPGFAVEKETEGLWLLINKPLDRBEEIAKYELFGHAVSENGASVED	197		
DB	189	ETTVFVITGPGADGPPGFAVEKETEGLWLLINKPLDRBEEIAKYELFGHAVSENGASVED	248		
QY	198	PMVSIIVTDQNDHKKPKFTQDFRGSGLGVLPQSVNQVATATEDDAIYTYNGVAVSI	257		
DB	249	PMEIVTVTDQNDHKKPKFTQDFRGSGLGVLPQSVNQVATATEDDAIYTYNGVAVSI	308		
QY	258	HQEPKDPHLMFTIHRSTGTISVSSGLDRKVPXYTLTIQATDMGDGTTTAVAVSE	317		
DB	309	LSQDPQLPHKNMFTVNRDGTIVSVTSGLDRESYPTLVVQAAQLQEGSLTTAKAVIT	368		
QY	318	ILDANDNAPMPOQKYEAVHPVNAVCHVEQRLTWTDLDPNAPNATYLLIMGGDDGH-	376		
DB	369	VKDINDNAPMPOQKYEAVHPVNAVCHVEQRLTWTDLDPNAPNATYLLIMGGDDGH-	426		
QY	377	FIITHPESNQIGILTRKGLDPEAKNQHTLVYVNTNEAPFVKLPSTATVIVHVEDVNE	436		
DB	427	FVITDPTKNEGLXTAKGLDPEAKNQHTLVYVNTNEAPFVKLPSTATVIVHVEDVNE	486		
QY	437	APVFPVPSKVVEQSGIPTGEBVCVYTAEDDPK-ENQKISYRIURDPAGMLAMPDPSQV	495		
DB	487	APVFPVPAERVRVPEDPFGVGLBIASTAREPDTFMEQKITVRIWRDTANMLINPETGVI	546		
QY	496	TAVGLDREDEQFVRNNIVYVWVLAMDNGSPPTGTGTLTLLTLVDVNDHGVPEPPOITI	555		
DB	547	SIPAENDRSDSEHKVNSITVALLIATDQSGFIATGTGTLVLLVSDVNDNAPIPEPRNQF	606		
QY	556	CNQSPVRHVINTDKDLSPTSPFOAQTDDSDIYVTAEVNE-EGDTVTVLSLKFKLKQDT	614		
DB	607	CORNPRPHVITILDPLDPTSPFTAEALTHGASVNWITTEYNDAEQESLILQPKRDLIGE	666		
QY	615	YDVHLSLSHGKNEQLTVIRATVCDCHGHVETC-----PGPWKGGRTLPVLA	662		
DB	667	YKINLKLSDNQNDQVITLLEHVCDCEGVNVMKALISLEAGLQVPA-----ILGLGG	720		
QY	663	VLALLFLVLLVLLVKKRKEIPKELLLPEDDTRNDNVFYEGEGGEEDQDYITQLHRL	722		
DB	721	ILALLIILLLLFLRRRTVWKEPILPPDDDDTRNDNVFYDEEGGEEDQDFDLSQLHRL	780		
QY	723	EAPPEVVLNDVAPITIPTRVPRPAPNDPEIGNPIENKAANTDTPAPYDTLLVFDY	782		
DB	781	DARPEVI-ENDAVPTLMSMPQVRPAPNDPEIGNPIDENKAASDTPAPYDLSLVFDY	839		
QY	783	EGSGDAAISLSLTSASDQDDQDYVLNWSGRFVKLADMYGGGEDD	829		
DB	840	EGSGSEASLSLNSSESDDQDDQDYVLNWSGRFVKLADMYGGGEED	886		

RESULT 7

CADF_XENLA

RESIST 7

RESULT /
CADEF XENLA

```

FT CONFLICT 3 3 S -> G (IN REF. 2).
FT CONFLICT 112 112 I -> N (IN REF. 2).
FT CONFLICT 573 573 S -> P (IN REF. 2).
FT CONFLICT 864 864 S -> P (IN REF. 2).
SQ SEQUENCE 880 AA; 97650 MW; 66CCDD91566427D86 CRC64;

Query Match 54.4%; Score 2375; DB 1; Length 880;
Best Local Similarity 54.2%; Pred. No. 3,7e-140;
Matches 471; Conservative 128; Mismatches 214; Indels 56; Gaps 9;

QY 11 LLLQVWMLQCAASEPCRAVFEAEVLEAGGAEQEPQALGV-FMGCPGQEPALFSTD 69
Db 17 LCLQVVPINADVSGCKGPFSSAEYFVSNRELEGRKLGKVNSDCTTRKHGYDVG 76
QY 70 NDDFTV-----RNGETVQERS 86
Db 77 DSRPRVLPDGTVLKRVHKLHDKFTISTWDARGIKHSTNIAVASKRHSRSGEASRS 136
QY 87 LKERNPLKIPSKRI-IRRKRQWVAPISVPENGKPPORLNOLKSNKORDTKIFYSI 145
Db 137 ---KLPLVTPPTHGLKRRKRDWVPIPIKVSNERGPPFRRLVQIKSNKDRFNKYYSI 193
QY 146 TPGGADSPPGVFAVEKETGMLINKPLDREIEAKYELFGHVSNGASVEDPMNISIV 205
Db 194 TGGADNPPQGVFRIEWTGMLVTRDLREEDYKVLSSHAVSENGSPVEEPMETINV 253
QY 206 TDQNDHKPKTQDTFRGSVLGVLPGTSMQVATDEDDAIYTYNGVAYSIHQEPKDP 265
Db 254 IDQNDNRPKTDVFRGSVREGVQPGQVMAVSATDEDDNIDSLNGVLISILKQDPPEP 313
QY 266 HDLFTTHRSTGTSVSSGLDREKPEYTLTQATDMGDDGTTTAVAVETILDANDNA 325
Db 314 IPNLTFTNRETGVLSLGTGDRKPEYTLTQATDLEGAGLSVSGKALIIQITDANDNA 373
QY 326 PMFDQKYEAPVNAVGEVQRLTVDLDPANSPAWRATYLMGDDGDHFTIITHPS 385
Db 374 PIFDKTYTALVPENEGFEVQRLSVTDLDMPGTPAWQVYKIR-VNEGGFNITTDPS 432
QY 386 NOGLTTRKGLDFAKNQHTLYEVNTNEAPFLVLPKLTSTATIVVHVEDVNEAVFVPPSK 445
Db 433 NOGLTTRAKGLDFELRQYVQIVTNEAEPFVPLPTSTATIVVTVEDVNEAEPFVPAVS 492
QY 446 VVEQEGIPTGEPVYVTAEDDPKEN-QKISYRILRDPAGLWAMDPSGOVTAVGTLDR 504
Db 493 RVDVSEDLRGEKISLVAQDPKQIQKLSYFIGNDPASWLTWNKDNQVITNGNLDRE 552
QY 505 DEQVRNIIYEVWLVANDNGSPPTGTGTLTLLIDVNDHGPVPEPROITICNOSPVRHV 564
Db 553 SE-YVKXNTYVIMLVTDGVSVGTGTGTLILHLVDVNDNGVPVSPRVFTMCDQNPQV 611
QY 565 LNIITDKLSPTSFOALQTLDDSDIYVTAENBEGDTVLVLSKKFKQDTYDVHLSLSH 624
Db 612 LTISDADIPENTYKYVLSHSGDLTWKALSDSGTSMLSPTQQLKKGDSYIVVLSDA 671
QY 625 GNKEQLTVIRATVCDCHGVETCPGWKGGFIIP-----VLGAVALLFLLVLLVLRKK 680
Db 672 QNNPQLTWNATVCSCEGKAQKQELVGGFDLPITLIVLGSVALLLFLFLFLFLKRR 731
QY 681 RKIKEPILLPDDTRDNVYVYEGGGEEDQDYDITOLHREGLARPEVLRNDVAPTI 740
Db 732 KVKKEPILLPDDTRDNIFVYEGGGEEDQDYDLSQHLRGLSRDP-IMENDVVTLP 790
QY 741 TMYRPRANPDEIGNFIENLKAANTDPATPYDTLLVFDYSGSGDAASLSLTSAS 800
Db 791 APHYRPRFNPDEIGNFIDENLDAANDTAPPYDLSLVFDYSGSGEASLSLSSNS 850
QY 801 DQDQDYDLNENWGSFRFKLADMYGGGDD 829
Db 851 NDEHDYNYLSDWGRFRKRLADMYGGDDDE 879

```

```

ID CADB_XENLA STANDARD; PRT; 884 AA.
AC P33152;
DT 01-OCT-1993 (Rel. 27, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Blastomere-cadherin precursor (B-cadherin) (XBCad).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8335;
RN [1]
RP SEQUENCE FROM N.A. PubMed=7531482;
RX MEDLINE=95151580; PubMed=7531482;
RA Mueller H.-A.J., Kuehl M., Finnemann S., Schneider S.,
RA van der Poel S.Z., Hausen P., Wedlich D.;
RT "Xenopus cadherins: the maternal pool comprises distinguishable
RT members of the family.";
RL Mech. Dev. 47:213-223 (1994).
RN [2]
RP SEQUENCE OF 438-884 FROM N.A.
RX MEDLINE=92062581; PubMed=1840622;
RA Herzberg F., Wildermuth V., Wedlich D.;
RT "Expression of XBCad, a novel cadherin, during oogenesis and early
RT development of Xenopus.";
RL Mech. Dev. 35:33-42 (1991).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Expressed in pituitary gland, lung and kidney.
CC -!- DEVELOPMENTAL STAGE: During oogenesis and early development.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X78546; CAA55292.1; ALT_INIT.
CC EMBL; X63719; CAA45251.1; -.
CC HSP; P09803; 1SUH.
CC InterPro; IPR002126; Cadherin.
CC InterPro; IPR002233; Cadherin_C_term.
CC Pfam; PF00028; cadherin; 5.
CC Pfam; PF01049; Cadherin_C_term; 1.
CC PRINTS; P00205; CADHERIN.
CC SMART; SM00112; CA; 4.
CC PROSITE; PS00232; CADHERIN_1; 3.
CC PROSITE; PS0268; CADHERIN_2; 4.
CC Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
CC Signal.
CC SIGNAL 1 26 POTENTIAL.
CC PROPEP 27 157 POTENTIAL.
CC CHAIN 158 884 BLASTOMERE-CADHERIN.
CC DOMAIN 158 706 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 707 730 POTENTIAL.
CC DOMAIN 731 884 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 158 265 CADHERIN 1.
CC DOMAIN 266 378 CADHERIN 2.
CC DOMAIN 379 489 CADHERIN 3.
CC DOMAIN 490 595 CADHERIN 4.
CC DOMAIN 596 706 CADHERIN 5.
CC DOMAIN 828 852 SER-RICH.
CC DOMAIN 879 884 ASP/GLU-RICH (ACIDIC).
CC CARBOHYD 427 427 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 560 560 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 683 683 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CONFLICT 438 440 ILT -> NSA (IN REF. 2).

```

FT	CONFLICT	677	677	R -> Q (IN REF. 2).
FT	CONFLICT	787	787	V -> A (IN REF. 2).
FT	CONFLICT	820	820	D -> N (IN REF. 2).
FT	CONFLICT	857	857	D -> N (IN REF. 2).
FT	CONFLICT	863	863	D -> N (IN REF. 2).
FT	CONFLICT	882	882	MISSING (IN REF. 2).
SQ	SEQUENCE	884	AA; 97980 MW; 98865D7E1DCB75B	CRC64;
Query Match				
Best Local Similarity 54.3%; Score 2371; DB 1; Length 884;				
Matches 469; Conservative 125; Mismatches 222; Indels 51; Gaps 8;				
Qy	11	LLLLQVCMQCAASEPCRAVPREAEVTLLEAGGAEQEPQALGV-FMGCPGQEPALFSTD	69	
Db	18	LCLLQVPSINVDSGQCPQSSANYTFSVNRLELRGKLGKLVDCVTRKHGLYDVG	77	
Qy	70	NDDFTVRNGTV-----QERRSLK	88	
Db	78	DSRFVLPDGTVLVGRHVKLHSLKSDTRFTISTWDARGIKHSTNISVVMKRHSRGEARSRS	137	
Qy	89	ERNPLKIPPSKRI-LRRHKDDWVAPISVPENKGPFPQRLNQLKSKNDRDTKIFYSITG	147	
Db	138	SELPLVTFPEKHGLKKEKRDWIPPIKVSNERGPPKRLVQIKSNKEKLSKVFSITG	197	
Qy	148	PGADSPPEGVFAVEKETGWLKLLNKPLDREBIAKYELFGHAVSENGASVEDPMNISIIVTD	207	
Db	198	QGADTPPEGIFRIKETGHWQVTEPLDREBEYKVVLLSHAVSENGASVEEPMETVITVID	257	
Qy	208	QNDHKPFTQDTRGSLVGLVPGTSVMQVATDEDDAIYVNGVAYSIHQSPKDPHD	267	
Db	258	QDNRPKFTQVPRGSRVREGVGTQKMSVSAATDDSDLSLNGVAYSILKQDPEPIP	317	
Qy	268	LMFTIRSTGTISVISGLDREKVPEYTLTIQATMDGDSGTTTAVAVVELLDANDANPM	327	
Db	318	NLFTINRETGVISLIGTGLDREKPEYTLTVQAADLDGAGLTAEGKAVIEITDANDNAPI	377	
Qy	328	FDQKYEAVHPENAVGHEVQRLVTDLDAFNSPAWRATYILMGDDGDHFTITTHPESNQ	387	
Db	378	FDEKTYTALVPENVEGFEVQRLSVTLDMPTAAQAVYKIR-VNEGPFENITTDPSNQ	436	
Qy	388	GILTRKGLDFAKNQHTLYVEVTNEAPFVLKPTSTATVHVVEDVNEAPFVPPSKV	447	
Db	437	GILTTAKGLDFEVRKQVVIQITVENAVPSPVLPSTATVTVTVEDVNEAPFVPPSVRV	496	
Qy	448	EVOEGITGTPVCVYTAEDPKEN-QKISYRIILDPAGWLAMPDPSGQVAVGTLDREDE	506	
Db	497	DVSEDLTRGEKIVSLVAQDPKQIQKLSYFIGNDPAWLTINKNGIVTNGNLDRESE	556	
Qy	507	QFVRNIYEVNVLAMNGSPPTGTGTLTLIDVNDHGVPEPRQITICNOSPVRHVLN	566	
Db	557	-YVKNNTYVIMLVTDGVPVGTGTGLILHLVLDINDNGFVPSFVFTMCDQNPQVLT	615	
Qy	567	ITDKLSPTSPPQAOLTDSDIYVTAENVNEEDTVVLSLKFLKQDHYDHLSDHGN	626	
Db	616	ITDADIPNTYFVSLSHSGELTWAEKLSKTSNLSFTQOLKGDYSIYVLLADAQA	675	
Qy	627	KEQLTVIRATVCDCHGVETCPQWKGGFILP-----VLGAVLALFLLLVLLLVKRRK	682	
Db	676	NRQLTVNATVCLCEGKAICQEKLVAGFDPILVILGSLAILLTLILLFLKRRKV	735	
Qy	683	IKEPILLPEDDTRNDYFYEGEGGEDDYDTOLHRLGLEAPVPEVVLNDVAPLIPTP	742	
Db	736	VKEPILLPEDDTRNFIYFEGEGGEDDYDLSQLHRLGLDAPD-IMRNDVWPTLMSVP	794	
Qy	743	MYRFRANPDEIGNFIENLKAANTDPTAPPYDILLVFDYSGSGDAASLSLTSASDQ	802	
Db	795	HYRFRSPNDEIGNFIDENLDAANDPTAPPYDLSLVFDYSGSGEAAASLSLNSNSN	854	
Qy	803	DQDYDVLNWSGFRKFLADMYGGGEDD	829	
Db	855	EHDYNYLNDWGRFRKLADMYGGDDDD	881	

RESULT 9				
CAD3_BOVIN	STANDARD; PRT; 491 AA.			
ID_CAD3_BOVIN	PI9535;			
AC	01-FEB-1991 (Rel. 17, Created)			
DT	01-FEB-1991 (Rel. 17, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Cadherin-3 (Placental-cadherin) (P-cadherin) (Fragment).			
GN	CDH3 OR CDHP.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=90360379; PubMed=2390969;			
RA	Liu C.W., Cannon C., Power M.D., Kiboneka P.K., Rubin L.L.;			
RT	"Identification and cloning of two species of cadherins in bovine			
RT	endothelial cells.";			
RL	EMBO J. 9:2701-2708 (1990).			
CC	-!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.			
CC	They preferentially interact with themselves in a homophilic			
CC	manner in connecting cells; cadherins may thus contribute to the			
CC	sorting of heterogeneous cell types.			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-!- SIMILARITY: Contains 5 cadherin domains.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC				
DR	EMBL; X53614; CAA37676.1; -.			
DR	FIR; S11694; IJBOCP.			
DR	InterPro; IPR002126; Cadherin.			
DR	InterPro; IPR00233; Cadherin_C term.			
DR	Pfam; PF00028; cadherin; 3.			
DR	Pfam; PF01049; Cadherin_C term; 1.			
DR	PRINTS; PR00205; CADHERIN.			
DR	SMART; SM00112; CA; 2.			
DR	PROSITE; PS00232; CADHERIN 1; 1.			
DR	PROSITE; PS0268; CADHERIN 2; 2.			
KW	Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.			
FT	NON_TER 1			
FT	DOMAIN <1 316 EXTRACELLULAR (POTENTIAL).			
FT	POTENTIAL 317 339			
FT	TRANSMEM 340 491			
FT	DOMAIN 347 491 CYTOPLASMIC (POTENTIAL).			
FT	DOMAIN <1 102			
FT	CADHERIN 3.			
FT	DOMAIN 103 208			
FT	CADHERIN 4.			
FT	DOMAIN 209 314			
FT	CADHERIN 5.			
FT	DOMAIN 447 462			
FT	SER-RICH.			
FT	CARBOHYD 228 228			
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).			
SQ	SEQUENCE 491 AA; 54207 MW; 65B0AE5918C4771 CRC64;			
Query Match				
Best Local Similarity 53.2%; Score 2323; DB 1; Length 491;				
Matches 434; Conservative 23; Mismatches 137; Indels 0; Gaps 0;				
Qy	339	ENAVGHEVQRLVTDLDAFNSPAWRATYILMGDDGDHFTITTHPESNGILTTTRKGLDF	398	
Db	1	ENTVSHEVQRLVTDLDAFNSPAWRATYIRIVGDNGDHFTITTDPSNGILTTQKGLDF	60	
Qy	399	EAKNQHTLYVEVTNEAPFVLKPTSTATVHVVEDVNEAPFVPPSKVVEGIPTEGP	458	
Db	61	EAKTQHTLYVEVINEPFFVKLPTSTATVTVVLEDVNEPFFVPPSKVIEIGISTGEP	120	
Qy	459	VCVTAEDPKENOKISYRIILDPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVNV	518	
Db	121	ICATYARDPKGSQKISYRIILDPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVNV	180	

RESULT 9	CAD3_BOVIN	STANDARD;	PRT;	491 AA.
ID	CAD3_BOVIN			
AC	P19535;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-FEB-1991 (Rel. 17, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Cadherin-3 (Placental-cadherin) (P-cadherin) (Fragment).			
GN	CDH3 OR CDHP.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Rutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=90360979; PubMed=2390969;			
RA	Liaw C.W., Cannon C., Power M.D., Kiboneta P.K., Rubin L.L.;			
RT	"Identification and cloning of two species of cadherins in bovine endothelial cells.";			
RL	EMBO J. 9:2701-2708(1990).			
CC	FUNCTION: Cadherins are calcium dependent cell adhesion proteins. They preferentially interact with themselves in a homophilic manner in connecting cells; cadherins may thus contribute to the sorting of heterogeneous cell types.			
CC	SUBCELLULAR LOCATION: Type I membrane protein.			
CC	SIMILARITY: Contains 5 cadherin domains.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
DR	EMBL; X53614; CAA37676.1; -			
DR	PIR; S11694; IJBOCP.			
DR	InterPro; IPR002126; Cadherin.			
DR	InterPro; IPR000233; Cadherin_C_term.			
DR	Pfam; PF00028; cadherin; 3.			
DR	Pfam; PF01049; Cadherin_C_term; 1.			
DR	PRINTS; PR00205; CADHERIN.			
DR	SMART; SM00112; CA; 2.			
DR	PROSITE; PS00232; CADHERIN_1; 1.			
DR	PROSITE; PS00268; CADHERIN_2; 2.			
KW	Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.			
FT	NON_TER 1			
FT	DOMAIN <1 316			EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 317 339			POTENTIAL.
FT	DOMAIN 340 491			CYTOPLASMIC (POTENTIAL).
FT	DOMAIN <1 102			CADHERIN 3.
FT	DOMAIN 103 208			CADHERIN 4.
FT	DOMAIN 209 314			CADHERIN 5.
FT	DOMAIN 447 462			SER-RICH.
FT	CARBOHYD 228 228			N-LINKED (GLCNAC. .) (POTENTIAL).
SQ	SEQUENCE 491 AA; 54207 MW; 68EB0AE5918C4771 CRC64;			
Query Match				
Best Local Similarity 53.2%; Score 2323; DB 1; Length 491;				
Matches 434; Conservative 23; Mismatches 34; Indels 0; Gaps 0;				
Qy	339	ENAVGHEVQRLVTDLDAFNSPAWRATYILMGDDGDHFTITTHPESNQILTRKGLDF	398	
Db	1	ENTVSHEVQRLVTDLDAFNSPAWRATYIRIVGGDNGDHFITITTPESNQILTRKGLDF	60	
Qy	399	EAKQHTLVYVTVNEAPVFLKPTSTATVHVVEDVNEAPVFPVPSKVVEQEGIPTCEP	458	
Db	61	EAKQHTLVYVTVNEAPVFLKPTSTATVHVVEDVNEAPVFPVPSKVVEQEGISTCEP	120	
Qy	459	VCVTAEDPDKENKISYRIILDPAGWLAMPDQSGQVAVGTLDREDEQFVRNIYEVW	518	
Db	121	ICAVTARPDKSGQKISVHILRDPAGWLAMPDQSGQVTAAGVLDRDEQFVRNIYEVW	180	

QY 519 LAMNGSPPTGTTGTTLLTLIDVNDHGPVPEPRQITICNQSPVRHVNLITDKLSPHTSP 578
Db 181 LATDGSPTTGTGTTLLTMDINDHGPVPEPRQITICNQSPVQVNLITDKLSPHTAP 240
QY 579 FOAQLTDSDIYTAENWEGDVTWLSLKFPLKQDYDVHLSLSDGNKEQLTVIRATVC 538
Db 241 FOAQLTHSDSVYTAENYKGDVALSLKFLKQGEYDVHLSLSDGNKEQLTVIRATVC 300
QY 639 DCHGHVETCPGPKGGPILVGLAVLALLFLLLVLLVLRKKGKIKPEPLLPEDDTRDNV 698
Db 301 DCHGNWVTCDPWTGFWPLILGNALLVLLVFLVLRKKGKIKPEPLLPEDDTRDNV 360
QY 699 FYEGEGGEGEDQDYDTQLHRLGLEARPEVVLVNDVAPTIPTPMYRPRPANDPDEIGNFI 758
Db 361 FYEGEGGEGEDQDYDTQLHRLGLEARPEVVLVNDVAPSIPTPMYRPRPANDPDEIGNFI 420
QY 759 IENLKAANTDPTAPPDYTLVDFYEGSGSDAASLSLTSSASDQDDYDYNLWNGSRFKK 818
Db 421 IENLKAANTDPTAPPDYTLVDFYEGSGSDAASLSLTSSASDQDDYDYNLWNGSRFKK 480
QY 819 LADMYGGGEGEDD 829
Db 481 LADMYGGGEGEDD 491

RESULT 10
CADI CHICK
ID CADI CHICK STANDARD; PRT; 887 AA.
AC P08641;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epithelial-cadherin precursor (E-cadherin) (Cadherin-1) (Liver cell
DE adhesion molecule) (L-CAM).
GN CDH1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_taxid=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89017248; PubMed=3174655;
RA Sorkin B.C., Hemperly J.J., Edelman G.M., Cunningham B.A.;
RT "Structure of the gene for the liver cell adhesion molecule, L-CAM.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:7617-7621(1988).
RN [2]
RP SEQUENCE OF 51-887 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=87204217; PubMed=3472238;
RA Gallin W.J., Sorkin B.C., Edelman G.M., Cunningham B.A.;
RT "Sequence analysis of a cDNA clone encoding the liver cell adhesion
RT molecule, L-CAM".
RL Proc. Natl. Acad. Sci. U.S.A. 84:2808-2812(1987).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types. E-cadherin is a ligand for
CC integrin alpha-5/beta-7.
CC -!- SUBUNIT: Homodimer
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Non-neural epithelial tissues.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M22190; AAA82572.1; -.

DR EMBL; M22180; AAA82572.1; JOINED.
DR EMBL; M22181; AAA82572.1; JOINED.
DR EMBL; M22182; AAA82572.1; JOINED.
DR EMBL; M22186; AAA82572.1; JOINED.
DR EMBL; M22183; AAA82572.1; JOINED.
DR EMBL; M22194; AAA82572.1; JOINED.
DR EMBL; M22184; AAA82572.1; JOINED.
DR EMBL; M22185; AAA82572.1; JOINED.
DR EMBL; M22189; AAA82572.1; JOINED.
DR EMBL; M22193; AAA82572.1; JOINED.
DR EMBL; M22187; AAA82572.1; JOINED.
DR EMBL; M22192; AAA82572.1; JOINED.
DR EMBL; M22191; AAA82572.1; JOINED.
DR EMBL; M22195; AAA82572.1; JOINED.
DR EMBL; M16260; AAA82573.1; -.
DR PIR; A30201; IJCHCL.
DR HSSP; P09803; ISUH.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000233; Cadherin_C-term.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin_C-term; 1.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW Signal.
FT SIGNAL 1 26 POTENTIAL.
FT PROPEP 27 160
FT CHAIN 161 887 EPIITHELIAL-CADHERIN.
FT DOMAIN 161 714 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 715 735 POTENTIAL.
FT DOMAIN 736 887 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 161 268 CADHERIN 1.
FT DOMAIN 269 381 CADHERIN 2.
FT DOMAIN 382 493 CADHERIN 3.
FT DOMAIN 494 599 CADHERIN 4.
FT DOMAIN 600 704 CADHERIN 5.
FT DOMAIN 844 855 SER-RICH.
FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .).
FT CARBOHYD 346 346 N-LINKED (GLCNAC. . .).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .).
FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .).
FT CONFLICT 140 140 T -> M (IN REF. 1).
FT SEQUENCE 887 AA; 97783 MW; F56ABA9779A94A40 CRC64;
Query Match 52.0%; Score 2274; DB 1; Length 887;
Best Local Similarity 52.7%; Pred. No. 7.2e-134;
Matches 461; Conservative 122; Mismatches 236; Indels 55; Gaps 13;
QY 8 LASLILLQVWLCQAASEPCRAVFAEVTLBAGAEQEPQALGKV-FWGCPGGEALF 66
Db 16 LVLLILLQVCCRCDAAAPCGPFAAETFSVFPQDSVAAGRELGRVSVFAACSGRPWAV 75
QY 67 -STD-----NDD-----FTVRNGETVQERRS----- 86
Db 76 VPTDFRKVNGDVGVSTKRLTLGRKISFTIYAQDAMGRHSARVTVGRHRRHHHHH 135
QY 87 -LKENPKLI-PP--SKRLRRHKKDVTVAISVPENKGFPPQELNOLKKNKDRDKIF 142
Db 136 HLQDTTPAVLTPPKHDPGFLRRQKEDWIPPIISLENHRGYPYPMRLVQIKSNKDKESKV 195
QY 143 YSITGPGADSPEGVFAVEKKTGMLLNKPLDRBIAKYLEFGHAVSENGASVEDPMNIS 202
Db 196 YSITQGGADSPVGVFIETRETGMLEVTQDLREKIDRYLLSHAVSAGQVDPDPWEII 255
QY 203 IIVTDQNDHKPKFTQDTFRGSLVGLVPGTSMQVMTATDEDDAIYTVNGVAYIHSQEP 262
Db 256 IIVTDQNDKPKVFIKEVFGVIEENAKPGTSVMTVNTATDADDVAVTNGIVSYISVQQP 315
QY 263 KPHDLMFTIHRSTGTISVSGDLREKVPETLTIQATMDGDGSGSTTAVAVVEILDAN 322
Db 316 PRPHQPMFTIDPAKGIISVLGTGLDRETTNTVLLIQATDQEGKLSNTAILEVTDA 375

Qy	323	DNAPEPQKYEAVPENAVGHEVORLTVTDLDPNPSFAWRATYVILMGDDGDGHFTITTH	382	RP	SEQUENCE OF 149-169.
Ds	376	DNIFNFTYEGVEENKPGTEVARLTVTDQAPGSPAQVYVHKISGNLDGAFSIITD	435	RX	MEDLINE=91347911; PubMed=1879345;
Qy	383	PESNOGILITKGLDFAKNOHTLYVEVINEAPVFLKLPSTATIVVHVVEDVNEAPVEVP	442	RA	Angres B., Mueller A.H.J., Kellermann J., Hausen P.;
Ds	436	PSTNGLTKAKGLDYTKSRYDLVTVENKVPISVITLSTASVLTVDVNEPPVFP	495	RT	"Differential expression of two cadherins in <i>Xenopus laevis</i> .";
Qy	443	PSKVVEQEGIPTEPCVVTAEEDPKX-NOKISYRILRDPAGWLMDPDSGGQVAVGTL	501	RL	Development 111:829-844(1991).
Ds	496	PIKRVGVPEDLPVGGQVTSYTAQDPDRMKITRYMGSDPAGWLVIHPENGIVTATQPL	555	CC	!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
Qy	502	DRDEQEVNRNIVEMVLMANDNSPPTGTGTLTLLTLDVNDHGPVPEPQITICNOSPV	561	CC	They preferentially interact with themselves in a homophilic
Ds	556	DRESVHAI-NSTYKAILAVDNGIPDTGTGTLTLLQDVNDNGPPEPRSFICRQPE	614	CC	manner in connecting cells; cadherins may thus contribute to the
Qy	562	RHVLNITDKLSPTSFPQALTDSDIYTAEVNEEGDVTLSLKKFLKQDIDYVHLSL	621	CC	sorting of heterogeneous cell types. E-cadherin is a ligand for
Ds	615	KQILSIYDKLPHTYFPKAALEHSGSNNTVETIRGQ-DELAAGLKELEPGYINPVKL	673	CC	integrin alpha-E/beta-7.
Qy	622	SDHGNKEQTVIRATVCDCHGVETCP-CPWKGFP-ILPVLGAVLLELLVLL	675	CC	!- SUBUNIT: Homodimer.
Ds	674	TDSQGAQVTVQKRAQVCECGTAKCERRSYIVGGLGVPAILGILGAILALLLL	733	CC	!- TISSUE SPECIFICITY: Non-neural epithelial tissues.
Qy	676	LVRKKRKEPLLPEDDTRDNVFFYEGEGGEDQDYDTQLHRLGLEARPEVVLNDVA	735	CC	!- DEVELOPMENTAL STAGE: Appears in the embryonic ectoderm during
Ds	734	FARRKVEKEPLPPEDMDNVDYDEGGEEDQDYLSQLHGLDARPEVI-RNDVA	792	CC	gastrulation when epidermal differentiation commences and it
Qy	736	PTIIPMYPRANPDEIGNFIENKANTDPTAPPYDTLVFDYEGSGDAASLSL	795	CC	disappears from the neural plate area upon neural induction.
Ds	793	PPLMAAPQYPRANPDEIGNFIDENLKAADTPTAPPYDLSLVFDYEGSGEATSLSL	852	CC	!- SIMILARITY: Contains 5 cadherin domains.
Qy	796	TSASDQDQDYLVNENGSFVKLADMYGGGEDD	829	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	the European Bioinformatics Institute. There are no restrictions on its
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	use by non-profit institutions as long as its content is in no way
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	modified and this statement is not removed. Usage by and for commercial
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	or send an email to license@isb-sib.ch).
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	EMBL; U04708; AAA93116.1; -
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	EMBL; L29057; AAA61489.1; -
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	EMBL; X75454; CAA53206.1; -
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	HSP; P09803; ISUH.
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	InterPro; IPR002126; Cadherin.
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	InterPro; IPR002333; Cadherin_C term.
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	Pfam; PF00028; cadherin_5.
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	Pfam; PF01049; Cadherin_C term; 1.
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	PRINTS; PR00205; CADHERIN.
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	SMART; SM00112; CA; 4.
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	PROSITE; PS00232; CADHERIN_1; 3.
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	PROSITE; PS0268; CADHERIN_2; 4.
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	Signal.
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	1 25 POTENTIAL.
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	26 148
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CHAIN 149 872
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	DOMAIN 149 701
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	TRANSMEM 702 722
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	DOMAIN 723 872
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	DOMAIN 244 254
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	DOMAIN 358 368
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	DOMAIN 577 587
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	DOMAIN 710 721
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CARBOHYD 209 209
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CARBOHYD 456 456
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CARBOHYD 552 552
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CARBOHYD 631 631
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CARBOHYD 669 669
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CONFLICT 242 242
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CONFLICT 332 332
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CONFLICT 487 487
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CONFLICT 501 502
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CONFLICT 539 541
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CONFLICT 548 548
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CONFLICT 557 557
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CONFLICT 567 567
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CONFLICT 602 604
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CONFLICT 638 640
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CONFLICT 640 647
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CONFLICT 660 660
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CONFLICT 770 770
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CONFLICT 842 842
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	CONFLICT 870 871
Ds	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	SEQUENCE 872 AA; 96065 MW; 08637967EEFB4664 CRC64;
Qy	853	NSSASDQDQDYLVNENGNFKLAELYGGGEDD	886	CC	Query Match 50.3%; Score 2198; DB 1; Length 872;

RESULT 11
CADI_XENLA
ID CADI_XENLA STANDARD; PRT; 872 AA.
AC P30944; Q91709;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epithelial-cadherin precursor (E-cadherin) (Uvomorulin) (XTCAD-1).
OS *Xenopus laevis* (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; *Xenopus*.
CX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95324376; PubMed=7600966;
RA Levine E., Lee C.H., Kintner C., Gumbiner B.M.;
RT "Selective disruption of E-cadherin function in early *Xenopus* embryos
by a dominant negative mutant.";
RL Development 120:901-909(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94368839; PubMed=8066449;
RA Tooi O., Fujii G., Tashiro K., Shiohara K.;
RT "Molecular cloning of cDNA for XTCAD-1, a novel *Xenopus* cadherin, and
its expression in adult tissues and embryos of *Xenopus laevis*.";
RL Biochim. Biophys. Acta 1219:121-128(1994).
RN [3]
RP SEQUENCE OF 149-872 FROM N.A.
RX MEDLINE=94363396; PubMed=8081882;
RA Broders F., Girault J.M., Simonneau L., Thierry J.P.;
RT "Sequence and distribution of *Xenopus laevis* E-cadherin transcripts.";
RL Cell Adhes. Commun. 1:265-277(1993).
RN [4]


```
FT PROPEP 24 159
FT CHAIN 160 906
FT DOMAIN 160 724
FT TRANSMEM 725 746
FT DOMAIN 747 906
FT DOMAIN 160 267
FT DOMAIN 268 382
FT DOMAIN 383 497
FT DOMAIN 498 603
FT DOMAIN 604 714
FT DOMAIN 863 878
FT CARBOHYD 190 190
FT CARBOHYD 273 273
FT CARBOHYD 325 325
FT CARBOHYD 402 402
FT CARBOHYD 572 572
FT CARBOHYD 651 651
FT CARBOHYD 692 692
FT CONFLICT 12 12
FT CONFLICT 16 16
FT CONFLICT 196 196
FT CONFLICT 212 212
FT CONFLICT 357 357
FT CONFLICT 867 867
SQ SEQUENCE 906 AA; 99851 MW; 72DDC7B8B57C7AFC CRC64;

Query Match 39.2%; Score 1714; DB 1; Length 906;
Best Local Similarity 42.8%; Pred. No. 5.28-99;
Matches 391; Conservative 129; Mismatches 291; Indels 102; Gaps 27;

QY 6 GPLASLLQLVQLQCAASEP-----CRAVFREAEVLEAGAEQEPGQALGKV-FMGCP 59
DQ 6 GALATLPLLLALLQASVEASGETALCKTGPE-DVYSAVLSKDVHGQPLNKFSCN 64
QY 60 GQ-----BPALFSTNDND--FTVFN-----GETVQER----- 84
DQ 65 GKRKQVSESPADPKVDEGDMGVAVRSFPLSSSEHAKFLIYAQDQEKQWQVAVKLSL 124
QY 85 -----SLKERNPLK--IFP---SKRI--LRHRKRDVWVAPISVPENKGPFGQRLNQLK 132
DQ 125 PTLTEBSKSAEAEVEIVFPQFNSHGHQQRQKRWIPINLPENSRGPFQELVRI 184
QY 133 SNKDRDKYFISITGPGADSPGEGVFAVEKETGMLLANKPLDREELAKYELFGHVSNG 192
DQ 185 SDRDKNLSRYSVTGPQADQPTGIFINPISGQLSVTKPLDREQIARPHRAVDING 244
QY 193 ASVEDPMNISITVDNDHKPKQDTFRGSVLEGLVPGTSVMQVATDEDDAIYNGV 252
DQ 245 NQVENPDIIVNIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMVTVAIDADDP-NALNGM 303
QY 253 VAYSISQEPKPDHLMFTIHRSTGTISVSSGLDREKVPETLTITQATMDGDD--GST 309
DQ 304 LRYRIVSQAPSTPSPNNFTINNETGDIITVAAGLDREKVOQYTLITQATDMENPTYGLS 363
QY 310 TTAVAVVEILDANDNAPMFPQKYEAVHPENAVGHEVQRVLTVDLDAFNSPAWRATYLM 369
DQ 364 NTATAVITVDVNDNPEFTAMFTFYGEVPERVNDIIVANLTVTDKQDQPHTPANNAVYRIS 423
QY 370 GGDGDGDFHTITTHESNOGILITRKGLDLPFAKQHTLYVEVTEAPVVLK---PTSTAT 426
DQ 424 GGDPTGFACTDPSNDGLVTVVVKPIDFTENFVLTVAAENQVLAQKIQHPPOSTAT 483
QY 427 IVHVEDNEAPVFPSPKVEVQEGIPTEGPEVCVYTAEDPDK-ENQKISRYLRDPAGW 485
DQ 484 VSVTVIDVNDNPFAPNPKIIRQEBGLHAGTMTLTFTAQDPDRYMQQNIRYTKLSDPANW 543
QY 486 LAMDPDSGOYTAGTLDRDEQFVRNIVYEMVLMNDGSPPTGTGTLTLTLLIDVNDHG 545
DQ 544 LKIDPVNGQITIAVLDRSPN-VKNINYNATFLASDNGIPPMGSGTGIQIYLLDINDNA 602
QY 546 PVPEPQITICNSPVRHVLNIT--DKDLSFHTSPFQAQUTDSDIY---WTAEVNEEG 600
DQ 603 PQVLPQEAETC-STEDPENSINITALDVIDPAGPFADLPFLSPVTKENWII-TLNGD 660
```

```
QY 601 TVVLSLK-KFLQDQTDYDHLSDHGN--KEOLTIVIRATVCDCHGH-----VETCPGPMK 652
DQ 661 FAQLNLIKLEAGIYEVIIVIIIDSNPPKSNISIRVKVCQDSDNGDCTDVRIVGAGL 720
QY 653 GGFILPVLGAVLALLFLVLLV-----RKRKIKIPELLLPEDDTRDNVFFYG 702
DQ 721 G-----TGAIIAIIILIIILILVLMFVVMKRRDKERQAKQLLIDPDDVRDNLKYD 774
QY 703 BEGGGEDODYDITQLHRLGLEAPEV-----VLNDVAPTITPTMYRPRPA--NPDEIG 755
DQ 775 BEGGGEDQYDLSLQDPTVEPDAIKPVGIRMDERP-IHABPOYFVRSAPHPGDI 833
QY 756 NFIENLKAANTDPTAPPYDTLLVFDYEGSGSDAASLSLTSASDQDQDYDYLNEWGR 815
DQ 834 DFINEGLKAANDPTAPPYDLSLVFDYEGSGTAGSLSSLNSSSSSGGQDYDYLNDWGR 893
QY 816 FKXLDYMGGGED 828
DQ 894 FKXLDYMGGGDD 906

RESULT 13
CAD2_CHICK STANDARD; PRT; 912 AA.
AC P10288; Q90630;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neural-cadherin precursor (N-cadherin) (cadherin-2).
GN CDH2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88153917; PubMed=2831236;
RA Hattori K., Nose A., Nagatuchi A., Takeichi M.;
RT "Cloning and expression of cDNA encoding a neural calcium-dependent
cell adhesion molecule: its identity in the cadherin gene family.";
RL J. Cell Biol. 106:873-881(1988).
RN [2]
RP SEQUENCE OF 1-25 FROM N.A.
RC STRAIN=Cornish white rock Cockerel;
RX MEDLINE=97354288; PubMed=9210582;
RA Li B., Paradies N.E., Brackenbury R.W.;
RT "Isolation and characterization of the promoter region of the chicken
N-cadherin gene.";
RL Gene 191:7-13(1997).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
They preferentially interact with themselves in a homophilic
manner in connecting cells; cadherins may thus contribute to the
sorting of heterogeneous cell types. N-cadherin may be involved in
neural recognition mechanism.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; X07277; CAA30258.1; -
CC EMBL; U15563; AAB62980.1; -
CC FIR; A29964; IJCHCN.
CC HSP; P15116; 1NCJ.
CC InterPro; IPR002126; Cadherin.
CC InterPro; IPR002333; Cadherin_C_term.
```

DR Pfam; PF00028; cadherin; 5.
 DR Pfam; PF01049; Cadherin C-term; 1.
 DR PRINTS; PR00205; CADHERIN.
 DR SMART; SMO0112; CA; 5.
 DR PROSITE; PS00232; CADHERIN 1; 3.
 DR PROSITE; PS0268; CADHERIN 2; 5.
 DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
 KW Signal.
 FT SIGNAL 1 28
 FT PROPEP 29 164
 FT CHAIN 165 912
 FT DOMAIN 165 729
 FT TRANSMEM 730 752
 FT DOMAIN 753 912
 FT DOMAIN 165 272
 FT DOMAIN 273 387
 FT DOMAIN 388 502
 FT DOMAIN 503 609
 FT DOMAIN 610 720
 FT DOMAIN 869 884
 FT CARBOHYD 278 278
 FT CARBOHYD 330 330
 FT CARBOHYD 407 407
 FT CARBOHYD 578 578
 FT CARBOHYD 628 628
 FT CARBOHYD 657 657
 FT CONFLICT 21 21
 FT SEQUENCE 912 AA; 100464 MW; 98ACAC9DC1FC489 CRC64;
 Query Match 39.0%; Score 1704.5; DB 1; Length 912;
 Best Local Similarity 42.6%; Pred. No. 2.1e-98;
 Matches 390; Conservative 132; Mismatches 284; Indels 109; Gaps 27;
 7 PLASILL-LQVCLQCAASEP-CRAVPRE-----AEVLEAGGAEQPGQALGV-FMGC 58
 14 PLALMLAALQAPKATCDMLCKMGFPEDVHSAVSRSHVH-----GQPLNVRFSQC 68
 59 PQQEPALF-STDNDDFTV-----RNGETVOERR----- 85
 69 DENRKIYGSSEPEDFRVEDGVVVAERSFOLSAEPTFVVSARDKETQEWQMKVLTLP 128
 86 -----SLKRNPLK--IFPSKI-----LRRHKRWVAVIPSPNGKPPQRLNOL 131
 129 EPAFTGASEKQKXIEDIFFWQYKDSHLLKROKRDWVIPPINLPNSRGPPFQELVRI 188
 132 KSNKDRDFTKIFVSTIGPADSPGPGVAVKETEGLMLLNKPLDREBIKAYELFCHAVSEN 191
 189 RSDRDKSLSLSVYTGFGADQPTGIFINISGLSVTKFELDREQIASFHLRAHADVNV 248
 192 GASVEDPMNISIIVTDQNDHKPKTQDTFRGSVLEGLVPGTSVMQVATDEDDAIYTYNG 251
 249 GNQVENPDIIVINIDMNDNRPFLHQVWNGTVPEGSKPGTYVMVTVAIDADDP-NAQNG 307
 252 VVAYSIHQEPKOPHDLMTHTSTGTSVTSGLDREKVPXYTLTITQATMDGD---GS 308
 308 MLYRILSQASPPSPNNFTINNETGDIITVAAGLDREKVQYTLIICATDMEGNPTYGL 367
 309 TTTAVAVYELDANAPMFPDQKYEAHVPAVNAVGHVQRLTVTDLDPNSPAWRAVYLI 368
 368 SNTATAVITVDVNDNRPPEFTAMTFYGEVPEVNRVDIVANLTVTDKQDQHTPAWNAVYQ 427
 369 MCGDDGDHFTITHPESNOGILTTKGLDFFKAKQHTLYVVTNEAPVPLK---PTSTA 425
 428 TGGDPTGQFTIITDPSNDGLVTVVVKPDIIDFTNMFVLTVAENQVPLAKGIOHPPOSTA 487
 426 TTVVHVEDVNEAPVFPSPKVEVQEGIPTGEPVCVYTAEDPDKENQKIS--YRILKDDPA 483
 488 TVSIITVDVNESGYFVVPKLVQRQEGGLAGLSMLTFTTARDPDVYMQOTSLRYSKLSDDPA 547
 484 GWLAMPDPSQVATVGTLDREDEQVRNIVYVMVLANMGSPBPTGTGLLILLIDVND 543
 548 NWLKIDPVGNGQITTTAVLDRE--SIYVQNNMNTATFLASDNGIPGSPGGLQIVLLDND 606

QY 544 HGPVPEPQITICNQSPVRHVNLIT--DKOLSPHTSPFQALTTDDSDIY---WTAEVNEE 598
 DB 607 NAPQNPKEATTC-ETLQPNAINITAVDPDIDPAGPFAFELPDPSPSIKENWTI-VRIS 664
 QY 599 GBTVVLSLK-KFLKQDVTYVHLSLSDHGN--KEQLTVIRATVCDCH-----GHVETCGP 650
 DB 665 GDHAQLSLRIRFLEAGIYDVPITVDSGNPHASSTSVLKVVCQCDINGDCTDVRIVGA 724
 QY 651 WKGGFILPVLGAVLALLFLVALLLV-----RKKRKIKEPLLLPDDTFRDNVVFY 700
 DB 725 GLG-----TGATITALLCIIILILVLMVFWVKRRDKERQAKQLLIDPEDVDRLNK 778
 QY 701 YGEGGGEEDQYDITQIHRLGLEARPEV-----VLRNDVAPTIPTMYRPPA--NPDE 753
 DB 779 YDEGGGEEDQYDLSQLQPDVTVEPAIKPVGIRRLDERP-IHAEPQYVRSAAHPGD 837
 QY 754 IGFNIENLKANTDPTAPPYDTLLVFDYEGSGDAASLSLTSSASDQDQDYDLNKGW 813
 DB 838 IGDFFNEGLKADNDPTAPPYDLSLLVFDYEGSGTAGLSLSLNSSSSGEGEQDYDLNDWG 897
 QY 814 SRFKKLADMYGGED 828
 DB 898 PRFKKLADMYGGDD 912
 RESULT 14
 CAD2_MOUSE STANDARD; PRT; 906 AA.
 ID_CAD2_MOUSE AC P15116; Q64260;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neural-cadherin precursor (N-cadherin) (Cadherin-2).
 GN CDH2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89346748; PubMed=2762814;
 RA Miyatani S., Shimamura K., Hatta M., Nagafuchi A., Nose A.,
 RA Matsunaga M., Hatta K., Takeichi M.;
 RT "Neural cadherin: role in selective cell-cell adhesion.";
 RL Science 245:631-635(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Tamura K.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 RP PARTIAL SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6;
 RX MEDLINE=92409532; PubMed=1528849;
 RA Miyatani S., Copeland N.G., Gilbert D.J., Jenkins N.A., Takeichi M.;
 RT "Genomic structure and chromosomal mapping of the mouse N-cadherin gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:8443-8447(1992).
 RN [4]
 RP DEVELOPMENTAL STAGE.
 RC STRAIN=C57BL/6; TISSUE=Testis;
 RX MEDLINE=97033837; PubMed=8879495;
 RA Munro S.B., Blaschuk O.W.;
 RT "A comprehensive survey of the cadherins expressed in the testes of fetal, immature, and adult mice utilizing the polymerase chain reaction.";
 RL Biol. Reprod. 55:822-827(1996).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 160-267.
 RX MEDLINE=95191680; PubMed=7885471;
 RA Shapiro L., Fannon A.M., Kwong P.D., Thompson A., Lehmann M.S.,
 RA Grubel G., Legrand J.-F., Als-Nielsen J., Colman D.R.,
 RA Hendrickson W.A.;
 RT "Structural basis of cell-cell adhesion by cadherins.";

RL Nature 374:327-337(1995).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (3.4 ANGSTROMS) OF 160-374.
 RX MEDLINE=98318235; PubMed=9655503;
 RA Tamura K., Shan W.S., Hendrickson W.A., Colman D.R., Shapiro L.;
 RT "Structure-function analysis of cell adhesion by neural (N-)
 cadherin.";
 RL Neuron 20:1153-1163(1998).
 CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
 CC They preferentially interact with themselves in a homophilic
 CC manner in connecting cells; cadherins may thus contribute to the
 CC sorting of heterogeneous cell types. N-cadherin may be involved in
 CC neuronal recognition mechanism.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DEVELOPMENTAL STAGE: Expressed at all stages of testicular
 CC development with highest levels found in testes of 21-day-old
 CC mice.
 CC -!- SIMILARITY: Contains 5 cadherin domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M31131; AAA37353.1; -;
 CC EMBL; AB008811; BAA23549.1; -;
 CC EMBL; S45011; AAB23556.1; -;
 CC PIR; A32759; IUMSCN.
 CC PDB; INCG; 10-JUL-95.
 CC PDB; INCH; 10-JUL-95.
 CC PDB; INCI; 10-JUL-95.
 CC PDB; INCU; 18-MAR-99.
 CC MGD; MGI:88355; Cdh2.
 CC GO; GO:0005912; C:adherens junction; IDA.
 CC InterPro; IPR002126; Cadherin.
 CC InterPro; IPR00233; Cadherin_C_term.
 CC Pfam; PF00038; cadherin; 5.
 CC Pfam; PF01049; Cadherin_C_term; 1.
 CC PRINTS; PR00205; CADHERIN.
 CC SMART; SM00112; CA; 5.
 CC PROSITE; PS00232; CADHERIN 1; 3.
 CC PROSITE; PS0268; CADHERIN 2; 5.
 CC Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
 KW Signal; 3D-structure.
 FT SIGNAL 1 23
 FT PROPEP 24 159
 FT CHAIN 160 906
 FT DOMAIN 160 724
 FT TRANSMEM 725 746
 FT DOMAIN 747 906
 FT DOMAIN 160 267
 FT DOMAIN 268 382
 FT DOMAIN 383 497
 FT DOMAIN 498 603
 FT DOMAIN 604 717
 FT DOMAIN 863 878
 FT CARBOHYD 190 190
 FT CARBOHYD 273 273
 FT CARBOHYD 325 325
 FT CARBOHYD 402 402
 FT CARBOHYD 572 572
 FT CARBOHYD 651 651
 FT CARBOHYD 692 692
 FT CONFLICT 7 9
 FT CONFLICT 565 565
 FT CONFLICT 567 567
 FT CONFLICT 624 624
 FT STRAND 166 169
 FT TURN 170 171
 FT STRAND 178 182

FT STRAND	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000
-----------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------

QY 704 EGGGEEDQDYITQHRGLRPEV-----VLKNDVAPTIPTMYRPRPA--NPDEIGN 756
 DB 776 EGGGEEDQDYLSLQCPDTPVDAIKPVGIRLDERP-IHAPFPQVPSAAPHGPDIGD 834
 QY 757 FIENLKAANDTPAPYDITLLVPDYGGSDAASLSLTSASDQDQDYDYLNEGSRF 816
 DB 835 FINEGLKAANDTPAPPYDLSLLVPDYGGSGTAGSLSSLSNSSSGGQDQDYDYLNDGPRF 894
 QY 817 KKLADMYGGGD 828
 DB 895 KKLADMYGGGD 906
 RESULT 15
 ID_CAD4 CHICK STANDARD; PRT; 913 AA.
 AC P24503.
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cadherin-4 precursor (Retinal-cadherin) (R-cadherin) (R-CAD).
 GN CDH4.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 SEQUENCE FROM N.A.
 RC TISSUE=Retina;
 RX MEDLINE=9129341; PubMed=1712604;
 RA Inuzuka H., Miyataki S., Takeichi M.;
 RT "R-cadherin: a novel Ca(2+)-dependent cell-cell adhesion molecule
 RT expressed in the retina.";
 RL Neuron 7:69-79 (1991).
 CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
 CC They preferentially interact with themselves in a homophilic
 CC manner in connecting cells; cadherins may thus contribute to the
 CC sorting of heterogeneous cell types. May play an important role in
 CC retinal development.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: Embryonic brain and neuronal retina.
 CC -!- DEVELOPMENTAL STAGE: Detected only after some degree of neuronal
 CC differentiation has taken place and persists at least up to the
 CC newly hatched stage.
 CC -!- SIMILARITY: Contains 5 cadherin domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D14459; BAA03356.1; .
 DR PIR; JH0424; IJCHCR.
 DR HSP; P15116; 1NCI.
 DR InterPro; IPR002126; Cadherin.
 DR InterPro; IPR002033; Cadherin_C_term.
 DR Pfam; PF00028; cadherin_5.
 DR Pfam; PF01049; Cadherin_C_term; 1.
 DR PRINTS; PR00205; CADHERIN.
 DR SMART; SM00112; CA; 5.
 DR PROSITE; PS00232; CADHERIN_1; 3.
 DR PROSITE; PS00268; CADHERIN_2; 5.
 DR Cell adhesion; Glycoprotein; transmembrane; Calcium-binding; Repeat;
 KW Signal.
 FT SIGNAL 1 19 POTENTIAL.
 FT PROPEP 20 166 POTENTIAL.
 FT CHAIN 167 913 CADHERIN-4.
 FT DOMAIN 167 731 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 732 753 POTENTIAL.
 FT DOMAIN 754 913 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 167 274 CADHERIN 1.
 FT DOMAIN 275 389 CADHERIN 2.
 FT DOMAIN 390 504 CADHERIN 3.
 FT DOMAIN 505 610 CADHERIN 4.
 FT DOMAIN 611 721 CADHERIN 5.
 FT DOMAIN 722 828 SER-RICH.
 FT CARBOHYD 829 885 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 886 906 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 907 928 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 929 950 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 951 972 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 973 994 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 995 1000 P -> T (IN ONE FORM).
 SQ SEQUENCE 913 AA; 100885 MW; BD2BB9EC815DD6 CRC64;
 Query Match 38.8%; Score 1694; DB 1; Length 913;
 Best Local Similarity 41.5%; Pred. No. 9.3e-98;
 Matches 383; Conservative 135; Mismatches 286; Indels 118; Gaps 25;
 QY 11 LLLQVWVLCQA-----SEPCRAVFEAEVLEAGAGAEQEGQALGV-FMGCPQGE 62
 DB 7 LLLVLLVWGSAAALNGDLTVRPTCKGFSEBDYTAFTVSNIME-CQKLLKVFNCAGNK 65
 QY 63 PALFSTNDNDFTVR----- 76
 DB 66 GRYETNSILDFKVRADGTMYAVHQQVMAKQILLMTAVTDPTLGRWEAIVRFLVGEKLQ 125
 QY 77 -NGETVQERRS-----LKERNPVKIPP-----SKILLRHKEDWVAPISVENGKGPPO 126
 DB 126 HNGHKPKGRKSGPVDLAQQSDTLTPMRQSHQSAKGLREQKDWIPPINVPENRGPPPO 185
 QY 127 RLNLQKSNKRDRTKIFYITGPAGDSPEGVFAVEKETGWLILNKLPLDREIAYELFGH 186
 DB 186 QLVIRSKDKKEIHRYSITGVAGDQPMVEFSDPVSGRMYVTFEPDREERASVHLRAH 245
 QY 187 AVSENGASVEDPMNISITVTONDHKFTQDTPRGSVLEGVLPQTSVQVATDEDDAI 246
 DB 246 AVDMGNKVENIDLYIYIDMNDNRPFINQVNGSVDEGSKFTYVMTVANDADST 305
 QY 247 YVINGVAYSITHSQPKDPLDMFTIHERSTGTISVSSGLDREKVPETLTIOATDMGD 306
 DB 306 -TANGKVRIRIVTQTPQSPSQMFTINSETGDIVTVAAGLDREKVOQVWV-VQATDMGN 364
 QY 307 ---GSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHVEVQRLVTVDLDAPNSPAWR 363
 DB 365 LNVGLSNTATAITITVDNDNPPEFTTSTYSGEVPEVNVVAVNLTVMDRQPHSPNN 424
 QY 364 AYILMGDDGDHFTITTHPSNQILITRKLDPKAKNQHTLVVEVNEAPVFLKLP-- 421
 DB 425 AYIRIISGDPGSHFTIRIDFTVNEGMYVWKAVIDEMMRAPMLTVMVSNQAPLASSIQMS 484
 QY 422 -TSTATIVVHVDVNEAPVFPVPPSVKVEVQEGIPTGEPVCVYTABDDPK-ENQKISYRIL 479
 DB 485 FQSTAGVTSIVTDVNEAPVFPVFNHKLIRLEGVPTGTLTTFSAVDPRFQOQAVRYSKL 544
 QY 480 RDPAGWLANDPSQGVAVTGLDREDEQFVRNIIYEVWVLANDNQSPPTGTGTLTLI 539
 DB 545 SDPANWLNINATNGQITTAADVLDRES-YIKNVVYEATFLAADNNGIPPSAGGTGLQIYLI 603
 QY 540 DVNDHGFVPEPQITCNQSPVRHVNIT--DKDLSPTSPPFQALTDSDIY---WTAE 594
 DB 604 DINDNAPELLPKAQIC-EKPNLVININATAADADIDNPVFPVFPVPSVSAVRKNWII- 661
 QY 595 VNEEGTVVLSLK-KFLKQDQTVHLSLSDHGNK--QLTVIRATVDC--HGHVETCPG 649
 DB 662 TRLNGVYQSLRIMYLEAGVYDVPIIVTDSGNPPFYNTSIKVKVCPDENGDCDTTGA 721
 QY 650 PWKGGFIPVLGAVLALLFLVLLV-----RKRKIKPELPPEDDTRDNV 699
 DB 722 VAAAGL---GTGATAILIICIIILLTWLLFVWVKRERKERTKQLLIDPEDDVRDNL 778

```

QY 700 YXEEGGGSEDDDYDITOLHGLEARPV-----VLNDVAPTITPTMTVRPRA 749
Db 779 KDEEGGSEDDDYDLSLQ-----QPTMDHVLNKAPGVRRVDERP-IGABEYPIRPV 832
QY 750 --NPDEIGNFIETENIKAANTPTAPPDTLLVFPYEGSGDAASLSLTSSASDDQDDY 807
Db 833 IHPGGDIGIFINEGRAADNDPTAPPYDLSLLVFPYEGSGTAGSVSLNSSSG-DQDYD 891
QY 808 YLNEWGSRFKKLADMYGGGEDD 829
Db 892 YLNDWGPRFKKLADMYGGGED 913

```

Search completed: September 21, 2004, 22:08:01
Job time : 20 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:02:53 : Search time 57 Seconds
(without alignments)
4588.852 Million cell updates/sec

Title: US-09-916-849A-1
Perfect score: 4369
Sequence: 1 MGLPRGFLASLLLVQVWLQ.....NEWSRFFKLADMYGGGEDD 829

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:**

- 1: sp_archaea:**
- 2: sp_bacteria:**
- 3: sp_fungi:**
- 4: sp_human:**
- 5: sp_invertebrate:**
- 6: sp_mammal:**
- 7: sp_muc:**
- 8: sp_organelle:**
- 9: sp_phase:**
- 10: sp_plant:**
- 11: sp_rodent:**
- 12: sp_virus:**
- 13: sp_vertebrate:**
- 14: sp_unclassified:**
- 15: sp_rvirus:**
- 16: sp_bacteriap:**
- 17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3564.5	81.6	822	11	Q8BSL6
2	3560	81.5	821	11	Q8BRE1
3	3433.5	78.6	753	11	Q80VY6
4	2400.5	54.9	901	4	Q9UII7
5	2216.5	50.7	821	4	Q9UII8
6	1893	43.3	864	13	Q9QZ37
7	1702	39.0	906	4	Q8N173
8	1701	38.9	906	11	Q8BS19
9	1638	37.5	824	4	Q8NB64
10	1578.5	36.1	922	13	P79883
11	1308	29.9	711	13	Q72VY8
12	1272.5	29.1	729	13	Q7SZW2
13	1245	28.5	677	13	Q7SZW3
14	1221.5	28.0	690	13	Q72VY6
15	1127	25.8	566	4	Q9HIG6
16	1033	23.6	714	11	Q8VDK4

ALIGNMENTS

RESULT 1

Q8BSL6 PRELIMINARY; PRT; 822 AA.
ID Q8BSL6;
AC Q8BSL6;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Cadherin 3.
GN CDH3.

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Forelimb;

RX MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";

RL Nature 420:563-573 (2002).

DR EMBL; AK031265; BAC27327.1; -.

DR MGD; MGI:88356; Cdh3.

DR GO; GO:0005886; C:plasma membrane; IDA.

DR InterPro; IPR002126; Cadherin.

DR InterPro; IPR000233; Cadherin_C_term.

DR Pfam; PF00028; cadherin; 5.

DR PRINTS; PRO0205; CADHERIN.

DR SMART; SM00112; CA; 4.

DR PROSITE; PS00232; CADHERIN_1; 3.

DR PROSITE; PS0268; CADHERIN_2; 4.

SQ SEQUENCE 822 AA; 90611 NW; 7C653D59210A595A CRC64;

Query Match

Best Local Similarity 81.6%; Score 3564.5; DB 11; Length 822;

Matches 681; Conservative 56; Mismatches 82; Indels 13; Gaps 5;

Q8Bgl1 mus musculu
Q8x490 rattus norv
Q8Bm92 mus musculu
Q86cz9 homo sapien
Q8ns22 homo sapien
Q9dfis1 xenopus lae
Q9dfis0 xenopus lae
Q8c7q6 mus musculu
Q77704 canis famil
Q80ws7 mus musculu
Q8n147 ciona intes
Q8ggh3 gallus gall
Q8vi68 mus musculu
Q7zyv7 gallus gall
Q93264 xenopus lae
Q86ud2 homo sapien
Q920m3 mus musculu
Q84k6 mus musculu
Q9hb01 homo sapien
Q9w6g5 brachydanio
Q9hb00 homo sapien
Q91838 xenopus lae
Q95yk9 ciona savi
Q8cb32 mus musculu
Q17281 botryllus s
Q86up1 homo sapien
Q8wnw5 sus scrofa
Q86up0 homo sapien
Q8ayd0 gallus gall

17 1031 23.6 714 11 Q8Bgl1
18 1021 23.4 714 11 Q8x490
19 1020.5 23.4 785 11 Q8Bm92
20 969.5 22.2 796 4 Q86CZ9
21 965.5 22.1 790 4 Q8NS22
22 953.5 21.8 792 13 Q9DFS1
23 953.5 21.8 792 13 Q9DFS0
24 951.5 21.8 796 11 Q8C7Q6
25 948.5 21.7 814 6 Q77704
26 944.5 21.6 788 11 Q80WS7
27 941.5 21.5 798 5 Q9NL47
28 939 21.5 798 13 Q8QGH3
29 937.5 21.5 788 11 Q8VI68
30 937 21.4 798 13 Q7ZIV7
31 935 21.4 794 13 Q93264
32 926.5 21.2 794 4 Q86UD2
33 921 21.1 801 11 Q9ZOM3
34 909 20.8 832 11 Q8C4K6
35 904.5 20.7 894 4 Q9HB01
36 903 20.7 490 13 Q9W6G5
37 898 20.6 840 4 Q9HB00
38 894 20.5 790 13 Q91838
39 874 20.0 910 5 Q95YK9
40 869.5 19.9 719 11 Q8CB32
41 837.5 19.2 906 5 Q17281
42 807.5 18.5 781 4 Q86UP1
43 805 18.4 782 6 Q8WNW5
44 792.5 18.1 819 4 Q86UP0
45 784.5 18.0 773 13 Q8AYD0

```

QY 1 MGLPRGFLA-SLLLOVCLWCAASEPCRAVF-REAETVLEAGGAEQPGQALGKVFMC 58
Db 1 MELLSPHAFLLLLQVCLWLSVSEPYRAGFGEAGVTLEVEGTDLEPSQVLGKVALAG 60
QY 59 PQEPALFSTNDFTVRNGETVQERRSLKERNPLKIPSKRILRRHRKRWVAPISVPE 118
Db 61 QGMEHA---DNGDIIMLRGTVOGKDHMS-----PPTRILRRRRKRWVMPPIFVPE 110
QY 119 NGKGPFPORLNQKSNKDRDTKIFYSITGPGADSPPEGVAEVEKGTGLLKNKPLDRREEI 178
Db 111 NGKGPFPORLNQKSNKDRGTKIFYSITGPGADSPPEGVTIEKSGWLLHMLDRREI 170
QY 179 AKYELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTPRGSVLEGVLPQTSVMQVT 238
Db 171 VKYELFGHAVSENGASVEEPMNISIIIVTDQNDKPKFTQDTPRGSVLEGVNPGTSVMQVT 230
QY 239 ATDEDDAIYVNGVAVSYTHSQEPKDPHDLMTFTHRSSTGTISVSSGLDRKVPYTLTI 298
Db 231 ATDEDDAVNTYNGVAVSYTHSQEPKDPHDLMTFTHKSTGTISVSSGLDRKVPYTLTV 290
QY 299 QATMDGSGSTTTAVAVVEILDANDNAPMFDPOKYBAHVPENAVHVEVQRLTVDLDAPN 358
Db 291 QATMDGSGSTTTAAEVVQILDANDNAPFEPOKYBAHVPENAVHVEVQRLTVDLDVFN 350
QY 359 SPAMRATYHILMGDDGDHFTITTHPESNOGILTRKGLDFAKNOHTLYEVVNEAPFVL 418
Db 351 SPAMRATYHILVGDDGDHFTITTHPESNOGILTRKGLDFAKNOHTLYEVVNEAPFV 410
QY 419 KLPTSTATIVHVEDVNEAPVFPVPSKVVEQSGIPTGEPVCVYTAEDPKENOKISYRI 478
Db 411 KLPTATATVAVHVKOVNEAPVFPVPSKVVEAQEGISIGELVCIVTAQDPKEDQKISYTI 470
QY 479 LRDPACGLAMDPSGQVAVTGLTDRDQFVRNNIYEVVWLAMDNGSPPTGTGTLTLL 538
Db 471 SRDPANLAVDPSGQITAGILDRDQFVKNNYEVVWLATDSGNPPTGTGTLTLL 530
QY 539 IDVNDHGPVPERQITCNQSPVRHVLNITDKLSPHTSFPQALTDSDIYWTAEVNEE 598
Db 531 TDINDHGPPEPRQIICNQSVPQVLNITDKLSPNSSFPQALTDSDIYWTAEVSEK 590
QY 599 GDTVLSLKFLKQDVTYDHLSDHGNKEQLTVIRATVCDCHGV-ETCPGPKGGFIL 657
Db 591 GDTVALSLKFLKQDVTYDHLSDHGNKEQLTVIRATVCDCHGVQVNDPFPKGGFIL 650
QY 658 PVLGAVLALLLLVLLVLRKRIKEPILLPEDDTRDNVYFYGGEGGEDDQYDITQ 717
Db 651 PILGAVLALLLLVLLVLRKRIKEPILLPEDDTRDNVYFYGGEGGEDDQYDITQ 710
QY 718 LHRGLEARPEVLRNDVAPTIIPTPMYRPRANPDEIGNFIENLKAANTDPTAPPYDTL 777
Db 711 LHRGLEARPEVLRNDVAPTIIPTPMYRPRANPDEIGNFIENLKAANTDPTAPPYDSL 770
QY 778 LVFDYEGSGSDAAS-SSITSASDQDQDYLYNEWGRFKKLADMYGGGEDD 829
Db 771 LVFDYEGSGSDAASLSLTSASDQDQDYLYNEWGRFKKLADMYGGGEDD 822

```

RESULT 2

```

Q8BRE1 PRELIMINARY; PRT; 821 AA.
ID Q8BRE1
AC Q8BRE1;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-NOV-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Cadherin 3.
GN CDH3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;

```

```

RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK045041; BAC32194.1; -.
DR MBL; MGI:88356; Cdh3.
DR GO; GO:0005886; C:plasma membrane; IDA.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR002333; Cadherin_C-term.
DR Pfam; PF00028; cadherin; 5.
DR PRINTS; PRO205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS00268; CADHERIN_2; 4.
DR PROSITE; PS00268; CADHERIN_2; 4.
SQ SEQUENCE 821 AA; 90512 MW; 168356CSFB5CCED0 CRC64;

Query Match 81.5%; Score 3560; DB 11; Length 821;
Best Local Similarity 81.6%; Pred. No. 6.7e-241;
Matches 679; Conservative 56; Mismatches 83; Indels 14; Gaps 5;

QY 1 MGLPRGFLA-SLLLOVCLWCAASEPCRAVF-REAETVLEAGGAEQPGQALGKVFMC 58
Db 1 MELLSPHAFLLLLQVCLWLSVSEPYRAGFGEAGVTLEVEGTDLEPSQVLGKVALAG- 59
QY 59 PQEPALFSTNDFTVRNGETVQERRSLKERNPLKIPSKRILRRHRKRWVAPISVPE 118
Db 60 ----QGMHADNGDIIMLRGTVOGKDHMS-----PPTRILRRRRKRWVMPPIFVPE 109
QY 119 NGKGPFPORLNQKSNKDRDTKIFYSITGPGADSPPEGVAEVEKGTGLLKNKPLDRREEI 178
Db 110 NGKGPFPORLNQKSNKDRGTKIFYSITGPGADSPPEGVTIEKSGWLLHMLDRREI 169
QY 179 AKYELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTPRGSVLEGVLPQTSVMQVT 238
Db 170 VKYELFGHAVSENGASVEEPMNISIIIVTDQNDKPKFTQDTPRGSVLEGVNPGTSVMQVT 229
QY 239 ATDEDDAIYVNGVAVSYTHSQEPKDPHDLMTFTHRSSTGTISVSSGLDRKVPYTLTI 298
Db 230 ATDEDDAVNTYNGVAVSYTHSQEPKDPHDLMTFTHKSTGTISVSSGLDRKVPYTLTV 289
QY 299 QATMDGSGSTTTAVAVVEILDANDNAPMFDPOKYBAHVPENAVHVEVQRLTVDLDAPN 358
Db 290 QATMDGSGSTTTAAEVVQILDANDNAPFEPOKYBAHVPENAVHVEVQRLTVDLDVFN 349
QY 359 SPAMRATYHILMGDDGDHFTITTHPESNOGILTRKGLDFAKNOHTLYEVVNEAPFVL 418
Db 350 SPAMRATYHILVGDDGDHFTITTHPESNOGILTRKGLDFAKNOHTLYEVVNEAPFV 409
QY 419 KLPTSTATIVHVEDVNEAPVFPVPSKVVEQSGIPTGEPVCVYTAEDPKENOKISYRI 478
Db 410 KLPTATATVAVHVKOVNEAPVFPVPSKVVEAQEGISIGELVCIVTAQDPKEDQKISYTI 469
QY 479 LRDPACGLAMDPSGQVAVTGLTDRDQFVRNNIYEVVWLAMDNGSPPTGTGTLTLL 538
Db 470 SRDPANLAVDPSGQITAGILDRDQFVKNNYEVVWLATDSGNPPTGTGTLTLL 529
QY 539 IDVNDHGPVPERQITCNQSPVRHVLNITDKLSPHTSFPQALTDSDIYWTAEVNEE 598
Db 530 TDINDHGPPEPRQIICNQSVPQVLNITDKLSPNSSFPQALTDSDIYWTAEVSEK 589
QY 599 GDTVLSLKFLKQDVTYDHLSDHGNKEQLTVIRATVCDCHGV-ETCPGPKGGFIL 657
Db 590 GDTVALSLKFLKQDVTYDHLSDHGNKEQLTVIRATVCDCHGVQVNDPFPKGGFIL 649
QY 658 PVLGAVLALLLLVLLVLRKRIKEPILLPEDDTRDNVYFYGGEGGEDDQYDITQ 717
Db 650 PILGAVLALLLLVLLVLRKRIKEPILLPEDDTRDNVYFYGGEGGEDDQYDITQ 709
QY 718 LHRGLEARPEVLRNDVAPTIIPTPMYRPRANPDEIGNFIENLKAANTDPTAPPYDTL 777

```


Db	80	VFTIKESGWLILHMLDREKIVKYLEYGHAVSNGASVEEPWNISIIITVDQNDKPKFT	139
Qy	217	QDTERGSLVGLPCTSNQVQVATDDEDAITYNGVWVAVSIHSGEPKDPKDHLMFTIHRS	276
Db	140	QDTERGSLVGLVGPVGTSNQVQVATDDEDAVNTYNGVWVAVSIHSGEPKDPKDHLMFTIHRS	199
Qy	277	GTISVISGLDREKVPYETLIQTATMDGSGSTTAVAVVEILDANDNAPMFPDQKYEAH	336
Db	200	GTISVISGLDREKVPYETLTVQATMDGSGSTTAAEAVVQIILDANDNAPEFEPQKYEAM	259
Qy	337	VPENAVGVHQLTVTDLDAENSPAWATYLIWGGDDGDHFTITTHPESNOGILITRKGL	396
Db	260	VPENAVGVHQLTVTDLDAENSPAWATYLIWGGDDGDHFTITTHPESNOGILITRKGL	319
Qy	397	DPEAKNQHTLYVEVNEAPFVVKLPSTATIVVHVEDVNEAPVFPVPSKVEVQEGIPGT	456
Db	320	DPEAKNQHTLYVEVNEAPFAVKLPATATVWVHVVDNEAPVFPVPSKVEAQEGISIG	379
Qy	457	EPVCYTTAEDPDKENOKISYRILBDPAGWLAMPDSGQVAVGTLDREDEQVFRNNIYEV	516
Db	380	ELVCYTTAEDPDKEDOKISYTSRDPANWLAVPDSGQITAAAGILDREDEQVFRNNIYEV	439
Qy	517	MYLMDNGSPPTTGTGLLTLTLDVNDHGPVPEPROITICNOSPVRHVLNITDKDLSPH	576
Db	440	MYLMDNGSPPTTGTGLLTLTLDVNDHGPVPEPROITICNOSPVRHVLNITDKDLSPH	499
Qy	577	SFPOQLTDDSDIYVWFAVNEBGTIVVLSKKZLKQDTVDVHLSVSDHGNKESQLTVIRAT	636
Db	500	SFPOQLTDDSDIYVWFAVNEBGTIVVLSKKZLKQDTVDVHLSVSDHGNKESQLTVIRAT	559
Qy	637	VCDCHGV-ETCPGPKWGGFTPLVGLGAVLALLFLVALLVLLVKKKKEIKPELILPEDDTR	695
Db	560	VCDCHGV-ETCPGPKWGGFTPLVGLGAVLALLFLVALLVLLVKKKKEIKPELILPEDDTR	619
Qy	696	DNVYFYGEGGSDQYDITQLHRLGLARPEVVLNENDVAPTIPTMYRPRPANDPEIG	755
Db	620	DNVYFYGEGGSDQYDITQLHRLGLARPEVVLNENDVAPTIPTMYRPRPANDPEIG	679
Qy	756	NFLIENKAANDTAPPYDITLAFVYEGSGSDAASLSLTSASDODODYVILNEWGSR	815
Db	680	NFLIENKAANDTAPPYDITLAFVYEGSGSDAASLSLTSASDODODYVILNEWGSR	739
Qy	816	FKKLADMYGGGEDD	829
Db	740	FKKLADMYGGGEDD	753
RESULT 4			
Q9UII7			
ID	Q9UII7	PRELIMINARY;	PRT; 901 AA.
AC	Q9UII7;		
DT	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	E-cadherin.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
ON	NCBI_TaxId=9606;		
OX	[1]		
RP	SEQUENCE FROM N.A.		
RA	Shibamoto S., Fukudome Y.;		
RL	"E-cadherin mutant.";		
CC	Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.		
CC	-!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.		
CC	THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC		
CC	MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE		
CC	SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).		
CC	-!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).		
DR	EMBL; AB025106; BAA8957.1; -.		
DR	HSSP; P09803; 1SUH.		
DR	GO; GO:0016020; C:membrane; IEA.		

308 DPFLPKMFTINRNTGTVISVVTGLDRESFPTVTLVQAAADLQEGSLSTATAVITVD 367
321 ANDNAPMEDPKQYEAHVAVENAVGVEQLTVDLDAFNSPAWATYILMGDDGDHHTIT 380
368 TNDNPFIFNP----- 377
381 THPESNGILATRKGLDFAKNQHTLYVEVNEAPFVLKLPSTATVWVHVEDVNEAPVF 440
378 -----TT--GLDFAEQYILHAVINNVVPEVSLTSTATVVDVLDVNEAPIF 425
441 VPPSKVVEVQGIPTGEFVCVYTAEDDPK-ENQKISYRILRDPAGWLAMPDSDQVTAVG 499
426 VPPEKRVESDFGQGBITSYTAQEPDTPMEQKITVIRIMRDTANWLEINPTGAISTRA 485
500 TLDREDEQFVNNIYVNVLMADNGSPPTGTGTLTLLTLDVNDHGVPEPRQITICNQS 559
486 ELDRDPFHVKNYSYTAIIATDNGSPVATGTGILLILSDVNDAIPPEPTIFFCERN 545
560 PVRHVLNTDKLSHTSPFOAQTDSDIYVTAEVNE-EGDTVVLSLKFLKQDQYDVH 618
546 PKPOVINIADLPNTSPFAELTHGASANNWITQYNDPTQESILPKMALEVDGKIN 605
619 LSLSDHNGKEOLTIVRAVTCCHGVETC--PGWKGF-----ILPVLGAVLALLFLLV 672
606 LKLDNQKQOVTTILEVSVCDCEGAAGVCRKAQVVEAGLQIPAILGILGGLLALLILL 665
673 LLLVVRKRIKEPLLLPEDTRDNVFFYBEGGGEEDQDYITQLHRLGLEARPEVVLRN 732
666 LLLLRRAVVKPELLPEDTRDNVFFYBEGGGEEDQDPLSQHLRLGLDAREVT-RN 724
733 DVAPTIPTMYRPRPNDPDEIGNFIENLKAANTDTPAPPYDILLVDFEGSGSDAASL 792
725 DVAPTLMSVPRYLPRPNDPDEIGNFIENLKAANTDTPAPPYDILLVDFEGSGSEASL 784
793 SSLTSSASDQDQDYDLNEMGSRFKLADMYGGGEDD 829
785 SSLNSESDDQDQDYDLNEMGNRRFKLADMYGGGEDD 821

RESULT 6
Q90237
ID Q90237 PRELIMINARY; PRT; 864 AA.
AC Q90237;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE E-cadherin.
GN CDH1.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21270060; PubMed=11376490;
RA Babb S.G., Barnett J., Doedens A.L., Cobb N., Liu Q., Sorkin B.C.,
PA Yelick P.C., Raymond P.A., Marrs J.A.;
RT "Zebrafish E-cadherin: expression during early embryogenesis and
RL regulation during brain development.";
RL Dev. Dyn. 221:231-237(2001).
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS: CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AF364811; AAK52054.1; -.
DR ZFIN; ZDB-GENE-010606-1; cdh1.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.

DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin_C_term; 1.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS00268; CADHERIN_2; 4.
KW Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
SQ SEQUENCE 864 AA; 95254 MW; 5ACA19538396397C CRC64;

Query Match 43.3%; Score 1893; DB 13; Length 864;
Best Local Similarity 47.0%; Pred. No. 6.5e-124;
Matches 394; Conservative 121; Mismatches 284; Indels 40; Gaps 9;

QY 27 CRAVFEAEVTLBAGABOEPQALGV-FMGCPQEPALFSTDNDDFVTRNGETVQER 85
DB 29 CTFGELEEFVFKVHRNHLHSGKRLGVTFNDCGRTLFQSIDKRFEINDGTVTLKR 88
QY 86 SLK-----ERPLKIFPSKRI-----LRSHKRDVWVAPIS 115
DB 89 QVTLHGHKVFVSHVANDSSGMKHTASVRVERVPAQVSSSDVDLTKNKKVKGWIIPIS 148
QY 116 VPENGKPPFRLNQLKSNKDRDTKIFYISITGPADSPPEGVFAVEKETGWLILNKLPLR 175
DB 149 VSENSKPPFMRLLVQIKSDYAIETRLAYKITGEGADLPKGIPTIDRLSGWVSVTQLDR 208
QY 176 ERIAKYELFGHAVSENGASVEDPMNLSITVTDNDHKPKFTQDTFRGSLVGLVPGTSYM 235
DB 209 EKASAKYLAHANGVADVTEKPMDIIVTVDNDKPKVFTQNPNGNVPEALEKEGVEFM 268
QY 236 QVTATDEDDAIYTYNGVAVSIHSEBPKDPHLMFTIHRSTGTSIVSSGLDREKPEYT 295
DB 269 TVTATDADDKENTDNADISVIISQDPPSPKPMFAINPVSGSIVLEKGLDREQWFRVT 328
QY 296 LTIQATDMGDGSTTAVAVVEILDANDNAPMFDPKYEAHVAVENAVGVEQLTVDLJD 355
DB 329 LVTTATDMNGEGLSTTGTAVITVTDSDNDNAPLFEQSSYTSAPENQVGVAVKLPTVDG 388
QY 356 APNSPAWATYILMGDDGDHFTITTHPESNGILATRKGLDFAKNQHTLYVEVNEAP 415
DB 389 EPESTAWSTKYQIAGDKGFFNISTGFSRLSGIITVPLDYKTKVILSVIVVNDK 448
QY 416 FVLKLPSTATVWVHVEDVNEAPFVLKLPSTATVWVHVEDVNEAPFVLKLPSTATVWVH 474
DB 449 FVGPLPTSTATVWVHVEDVNEAPFVLKLPSTATVWVHVEDVNEAPFVLKLPSTATVWVH 508
QY 475 SVRILRDPAGWLAMPDSDQVTAAGTGLDREDSQFVRNNIYVNVLMADNG-SPPTGTGT 533
DB 509 TTRIGNDPSDWLNI-TGSCQIQVKAALDRESN-VDGKIKALILALDNDVESPATGTGT 566
QY 534 LLLTLIDVNDHGVPEPRQITICNQSPPVRHVLNITDKLSHTSPFOAQLTDDSDIYVTA 593
DB 567 LVIELQDVNDNAPVINERTIKLCNRESAPVLLSITDKLPPFAGPFKVEPQDTSKNMSV 626
QY 594 EVNEEGDVTVLSLKFQDQDYDLNEMGSRFKLADMYGGGEDD 653
DB 627 FNETGHFNLIKPSQLEGGYKVVLRVADREGESQENIIQASVCDCEGAFQCTDKQVA 686
QY 654 GF-----ILPVLGAVLALLFLVLLVLRKRIKEPLLLPEDTRDNVFFYBEGGGE 709
DB 687 GIPLFGVLGVGLILLALLLALLLMLFRKNSKKEPLLPEDDVRDNIYYDESGGED 746
QY 710 DDYDITQLHRLGLEARPEVVLNDVAPTIPTMYRPRPNDPDEIGNFIENLKAANTDP 769
DB 747 DQDFLSVLHRLGNRPE-VFRNDVAPTFMPAPQYRPRPANPEIETGTFIDNKLKADNDP 805
QY 770 TAPPYDILLVDFEGSGSDAASLSTSSASDQDQDYDLNEMGSRFKLADMYGGGEDD 828
DB 806 TAPPYDILLVDFEGSGSDAASLSTSSASDQDQDYDLNEMGSRFKLADMYGGGEDD 864

RESULT 7

```

Q8N173
ID Q8N173 PRELIMINARY; PRT: 906 AA.
AC Q8N173;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cadherin 2, type 1, N-cadherin (Neuronal).
DS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; BC036470; AAH36470.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR002233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin_C_term; 1.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 5.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 3.
KW Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
SQ SEQUENCE 906 AA; 97774 MW; 468BAB2CDAAC5A9A CRC64;

Query Match 39.0%; Score 1702; DB 4; Length 906;
Best Local Similarity 42.7%; Pred. No. 1.8e-110;
Matches 390; Conservative 129; Mismatches 292; Indels 102; Gaps 27;

QY 6 GPLASLLQLVCWLOCASEP-----CRAVFEAEVTLLEAGGAQEPQALGKV-FWGCP 59
DB 6 GALTLLPLLAALQASVEASGETALCKTGPE-DVISAULSKDVHSGPLNVKFSNCN 64
QY 60 GQ-----PAPLSTDNDD--FTVRN-----GETVQER----- 84
DB 65 GKRKQVYESSEPADPKVDGDMGVAVRFPPLSBEAKFLIYAQDKETQEKWQAVKLSLK 124
QY 85 -----RSLKERNPLK--LPP-----SKRI--LRHKEDVWVAPISVPENKGPFPQRLNQLK 132
DB 125 PTLTEESVKESAEVEIYFPKQFSKSHGLQRKRDWISPINLPENSRGFPFQELVKIR 184
QY 133 SNKRDRTKIFYSITGCGADSPPEGVFAVEKETGTMLLNKPLDRBEEIAKYELFCHAVSENG 192
DB 195 SDRDNKLSRLSVTGFGADQPTGTGFIINPISGQLSVTKPLDRDQIAFHLRAHAVDING 244
QY 193 ASVEDPMNISIIVTQNDHKPKFTODTPRGVLEGLVPGTSVMQVATDEDAITYNGV 252
DB 245 NQVENPIDIVINVDNDRNPEFLHQVWNGTVPEGSKEGTVMVTVAIDADDP-NALNGM 303
QY 253 VAYSTHSQPKDPHDMFTTHSTGTISVISGLDRKVPVTLTIQATMDGD---GST 309
DB 304 LRIRVSQAPSTPSNMFTINNETGDIITVAAGLDREKQQYTLIIQATMEGNTYGLS 363
QY 310 TTAVAVVEILDANDAPMFPDQKYEAPHPENAVGHEVQRLTVTDLDAFNSPANKATYILM 369
DB 364 NTATAVITVDNDNPPPEFTAMTFYGEVPEENRVDIIIVANLTVTDKQDPTPAWNAVYRIS 423
QY 370 GGDGDGHFTIITHPESNOGILTRGLDPEAKNQHTLVVEVINEAPFVKL---PTSTAT 426
DB 424 GGDPTGRFAIQDQNSNDGLVTVVFPIDPEANRMTVLTVAAENQVPLAKGIQHPQSTAT 483

```

Query Match 38.9%; Score 1701; DB 11; Length 906;
 Best Local Similarity 42.6%; Pred. No. 2.1e-110;
 Matches 391; Conservative 126; Mismatches 293; Indels 108; Gaps 28;

```

QY 2 GLPRGLASILLQVCWLCQAASEP-----CRAVFEAEVTLLEAGGAEPQGGALGKV-F 55
Db 6 GAPR-----TLPLLAALLQASVETSGEIALCKTGFFE-DVYSAVLPKDVHGQPLLVKXF 60
QY 56 MGC-----PGQEPALFSTDND--FTVRN-----GETVQER----- 84
Db 61 SNCNRKRVQVESSEPADFKYDEGTVYAVRSFPLTAQAKFLIYAQKETQEKQVAVN 120
QY 85 -----RSLKERNPLK--IPFSK-----RI LRHKRDWVAPISVPENKGGKPPFQRL 128
Db 121 LSREFTLTEPMKEPHEIEEIVFRLQAKHSGALQKQKRDWVIPPINLPENSRGPPFQEL 180
QY 129 NOLKSNKDRDKIFVSTGPGADSPPEGVFAVEKETGWLILNKLPLDREETAKYELFGHAV 188
Db 181 VRIISDRXNLISYSVTGPGADPTGIFINIISQGLSVTKEDRELRLARFLRAHAV 240
QY 189 SENGASVEDPMNIIIVTDQNDHKPKFTQDTPFRGSVLGVLPGTSVMQVATDDEDAIYT 248
Db 241 DINGNOVENPDIIVINVDMDNEPEFLHQWNGSVPEGSKPGTYVMTVTAIDADDP-NA 299
QY 249 YNGVVAYSIHSQEPKDHLDLMTHTSTGTISVSSGLDREKYPEYTLTQATMDGCD-- 306
Db 300 LNMRLRILSQASTSPNNFTINNETGDIITVATGLDREKVOQYTLIQTADMEGNPT 359
QY 307 -GSTTTAVAVVEILDANDNAPMFPDQKYEAHVPENAVGVHEVQRLTVTDLDAPNSPAWRAT 365
Db 360 YGLSNTATAVITVDNDNRPETAMTFYGEVPENRVDVIVANITVTDKQPHTPAWNAA 419
QY 366 YLINGGDDGHFTTTHPESNOGLITRKGLDFAKNQHTLYVEVNEAPFVKL-----PT 422
Db 420 YRISGGDPTGFAITLDPNSDGLVTVVKPFDIFENRMLVLTVAANQVFLAKGIQHPQ 479
QY 423 STATIVVHVEDNEAPVFPVPSKVVEQGIPTCEPVCVVTAEDPDK-ENQKISYRLRD 481
Db 480 STAIVSVTIVDVENPFAFNPKIIRQEBGLHAGTMTLTITAQDPDRYMQNIRYKLS 539
QY 482 PAGWLAMPDSQGVAVTGLDREDEQVRNNIYEVWVLAMDNGSPPTGTGTLTLTLDV 541
Db 540 PANWLKIDPVNGQITTTAVLDRESNP-VKNNIYNATFLASDNGIIPMWSGTGLQIYLLDI 598
QY 542 NDHGPVPEPQITICNOSPVRHVLNIT--DKDLSPTSPFQAQLTDDSDIY---WTAENV 596
Db 599 NDNAPQVLPQEAETC-ETPEPNSINITALDYDIDPNAGPFAFDLPUSPTVKRNT--IN 655
QY 597 E-EGDTVVLISLK-KFLQDQTYDVHLSLSDHGN--KEQLTVIRATVCDCHGH-----VETC 647
Db 656 RLNGDFAQLMLKIFLEAGIVEVPIITDGSNPPKSNISILRVKVCQDSNGDCTVDRI 715
QY 648 PGWKGFIPLVIGAVLALLPLLLVLLLY-----RKRKIKEPILLPDEDDTRDN 697
Db 716 VGAGLG-----TGAIITAILLCIILLILVLMFYVMKRDKERQAKQLLIDPEDDVRN 769
QY 698 VFYEGEGGGEEDQDYDITOLHRLGLEARPEV-----VLRNDVAPTITPTMYRPERA--N 750
Db 770 ILKYDESGGGEEDQDYLSQLQPDVTEPDALXPVGIRLDERP-THAEQYVPRGAAPH 828
QY 751 PDEIGNFIENLKAANTDTPAPPYDTLLVFDYEGSGDAASLSLSSASDQDQDYDYL 810
Db 829 PGDIGDFINEGLKAANDTAPPYDLSLLVFDYEGSGTAGSLSSLSNSSSGGQDYDYL 888
QY 811 EWGSRFKKLADMYGGGBD 828
Db 889 DWGPRFKKLADMYGGGDD 906

RESULT 9
Q8NB64 PRELIMINARY; PRT; 824 AA.
AC Q8NB64
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ34177.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Houta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
RA Matanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
RA Ishii S., Yamanoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagaesuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AK091496; BAC03677.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin_C term; 1.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 5.
DR PROSITE; PS00322; CADHERIN_1; 3.
DR PROSITE; PS00268; CADHERIN_2; 3.
KW Hypothetical protein; Calcium; Calcium-binding; Cell adhesion;
KW Glycoprotein.
SQ SEQUENCE 824 AA; 90261 MW; 31A0C127BD40BAEB CRC64;

Query Match 37.5%; Score 1638; DB 4; Length 824;
Best Local Similarity 43.0%; Pred. No. 4.7e-106;
Matches 360; Conservative 135; Mismatches 271; Indels 72; Gaps 23;

QY 41 CGAEQEPGQALCKVWGCPGQBPALFSTDNDFTVRNGETVQERSLKERNP-----L 93
Db 10 GGREN-----AGVTSKSCVGTGQYETSSP-----HSGHKPQKGVKVALDSPPKDTL 60
QY 94 KIFPSKRI---LRRHKRDWVAPISVPENKGGKPPFQRLNQLKSNKDRDKIFYSITGPGA 150
Db 61 LWPQHQNANGLRRRRKRDWVIPPINVPENSRGPPFQQLVIRSKDNDIPRYSITGVA 120
QY 151 DSPPSGVFAVEKETGWLILNKLPLDREETAKYELFGHAVSENGASVEDPMNIIIVTDND 210
Db 121 DQPPMEVFSIDMSGRMTVTRPMDREEHASVHLRAHAVDMNGNKVENPIDIYIYVDND 180
QY 211 HKPKFTQDTPFRGSVLGVLPGTSVMQVATDDEDAIYTYNGVVAYSIHSQEPKDPHLMF 270
Db 181 NRPEINQVYNGSVDEGSKPGTYVMTVTAIDADDST-TANGMVRVRIYVITQTPQSQNW 239
QY 271 TIHRTGTISVSSGLDREKYPEYTLTQATMDGCD---GSTTTAVAVVEILDANDNAPM 327
Db 240 TINSETGDIIVTAAGLDREKVOQYTVIYQATDMEGNLNYLSNTATAIITVTDNDNPE 299
QY 328 FDPQKYEAHVPENAVGVHEVQRLTVTDLDAPNSPAWRATYILMGDDGHFTTTHPESNQ 387
Db 300 FTASTFAGEVPENRVETVAVNITVMDRDQPHSPKNNAVRIISGDPGSHFVTRDPTVNE 359
QY 388 GLLTTRKGLDFAKNQHTLYVEVNEAPFVKLKP---TSTATTIVVHVEDNEAPVFPVPS 444
Db 360 GMVTVVKAVDYELARAFMLTVMVSNQAPLASGIQMSFQSTAGVTISIMDINEAPVPSNH 419
QY 445 KVEVEQEGIPTGPEVVCVVTAEDEPK-ENQKISYRLRDPAGWLAMPDSQGVAVTGLTDR 503

```

```

Db 420 KLIRLEGGVPPGTVLTTFSAVDPRFQQAQVRYSKLSDFASWHLHINATNGQITTAVALDR 479
QY 504 EDEQFVNNIYEVNVLAMNGSPPTCTGTLTLLTLDVNDHGVPPBPQITTCNQSPVRH 563
Db 480 E-SLYTKNNVYEAFLAAGNIPASGTGLQIYLDINDNAPELLKPKAQIC-ERNLN 537
QY 564 VLNIT--DKDLSPTSFPQALTDSDIYMTAEVNE-----EGDTVWLSLK-KFLKQD 613
Db 538 AINITAADADVDPNIGFVVEP-----FVPAARVKNWTITRLNGDYAQLSLRLYLEAG 592
QY 614 TYDHLSDSHGNK--EQLTVIRATVDC--HGHVETCPGPWKGGFLLPVGLAVALLFL 669
Db 593 MYDPIIIVTDSGNPPLSNTSIKVKVPCDDNGDCTTIGAVAAAGL---GTGAIVALLIC 649
QY 670 LVLVLLV-----RKKRKIKPELLLPEDDTRDNVFFYEGEGGEDDYDITQLH 719
Db 650 ILIILLTWLLFVWMMKREKERTKQLLDPEDDVRNLIKYEDEGGGEDDYDLSQLQ 709
QY 720 RGLRARPV-----VLNRVAPTIITPMY--RPRPANTDEIGNFIENLKAANTDPTA 771
Db 710 QP-EAMGHVPSKAPGVRRVDRP-VGAEPQYPIRPMVPHFPCDIDGDFINEGLRAANDPTA 767
QY 772 PPTLLAVDFYEGSGDAASLSSTSSASDQDDYDVLNWSGRFKKLADMYGGGEDD 829
Db 768 PYSLLVDFYEGSGTAGSVSSLSNSSSG--DQDYDVLNDWGRFKKLADMYGGGEDD 824

```

RESULT 10

```

P79883
ID P79883 PRELIMINARY; PRT; 922 AA.
AC
DB 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DE Maternally expressed neural cadherin Xnm-cadherin.
GN MN-CADHERIN.
OS
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC xenopodinae.
OX NCBI_TaxID=8353;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Tail bud;
RX MEDLINE=96257964; PubMed=8652409; Koga C., Ito Y., Hikasa H.,
RA Tashiro K., Tooi O., Nakamura H.,
RA Shikawa K.;
RT "Cloning and expression studies of cDNA for a novel Xenopus cadherin
RT (Xnm-cadherin), expressed maternally and later neural-specifically in
RT embryogenesis.";
RL Mech. Dev. 54:161-171(1996).
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; S82457; AAB37685.2; -.
DR HSSP; P15116; 1NCI.
DR GO; GO:0046020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR PRINTS; PF01049; Cadherin_C_term; 1.
DR SMART; SM00112; CA; 5.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 3.
KW Calcium; Calcium-binding; Cell adhesion; Glycoprotein.

```

```

SQ SEQUENCE 922 AA; 101173 MW; FEA7A95CBAF1B640 CRC64;
Query Match 36.1%; Score 1578.5; DB 13; Length 922;
Best Local Similarity 41.4%; Pred. No. 8.4e-102;
Matches 364; Conservative 127; Mismatches 275; Indels 113; Gaps 25;
QY 48 GOALGKV-FMCC-PQEBPALFSTNDPFTVNGETVQERRSLK----- 88
Db 60 GQKLKVPFGCGSSGAQGIWYETNNPDKVGADGAVYTAREVOIPAKOAKFIVAANDHET 119
QY 89 -----ERNPLK-----IPPSK-----RIURHKRWVV 111
Db 120 PEKWEAAIQLFVEETPLNQSOVQTGSQSESEQSQSGTLLPWRQHHKGLRRQKRDWVI 179
QY 112 APISVPENKGGFPQRLNQLKSNKDRPKIFYSTITGCGADSPPEGVFAVEKETGMLLNK 171
Db 180 PPNVNPENSGRFFPQQLVLIIRSDKDRDITRISTITGVGADQPPMAIFNIDIFGRMNVTR 239
QY 172 PLDRBEIAKYELFGHVAENGASVEDPMNIIIVTDQNDHKPKFTQDTFRGSVLEGLVLP 231
Db 240 PLDRERSYHLRAHVDINGKVENPIDLSIYVIMDMNDRPEFSSPIFNGSVDEASKPG 299
QY 232 TSMQVATDEDDAIYTVNGVVAISHQEPKDPHDLNFTTHRSTGTISVSSGLDREKV 291
Db 300 TYVMTVTAHDADD-INTSGIWMYRIMDQSPQSPSHDMFVHSKTGTVINTVAAGLDREKV 358
QY 292 PLYTLTIQATMDGDP---GSTTTAVVVEILDANDNAPMDPQKYEAHVPENAVGEYQR 348
Db 359 QQTIVVIQATMEGNLHGLSNTATITVADVNDNPEFTKMFIGVPEHNVVVVAN 418
QY 349 LTVTDLAPNSPAWATYILMGDDGDHFTTTHPESNOGILTTTRKGLDFPAKQHTLYV 408
Db 419 LTVVDRDQPYTSNNNAVFKIISGDPDGHFTIKTDFVTNEGIVTSKPYDYEWSKVFLIV 478
QY 409 EYVNEAPFVLKLP---TSTATIVHVEDVNEAPVFPVPSKVVEQEGIPTGECVCTAE 465
Db 479 MYTNOAPLASIQMSLOSTAAVTVSVNDVNEAPFPNRPNEPIRKLEGSAGRLITFSAV 538
QY 466 DPKENQKI-SYRIILRDPAGLAMPDPSQVTVAGTLDREDEQFVRNNIYEVNVLAMNG 524
Db 539 DPHSNMQQVLRYSKISDPANLAINTTNGQVSTTAVLDRE-SPFVKDDLYQAKFLATDNG 597
QY 525 SPPTTGTGTLTLLTLDVNDHGVPPBPQITTCNQSPVRHVNIT--DKDLSPTSPPQAA 582
Db 598 NPPASGTGTLTLLQI:DIINDNAPELLKPKAQIC-ERNPENGINITDAIDVDRPSADPFVE 656
QY 583 LTDDSDIY-----WTA-EVNEEGDTVWLSLKFLKQDITYDHLSDSHGNKE--QLTVIR 634
Db 657 L--PSVPYTIERNWTTHRINSIYARLSLIQI-GYLESGMYDVFIIVTDSGNPPLVNTSIK 713
QY 635 ATVCDC--HGHVETCPGPWKGGFLLPVGLAVALLFLVLLVLLV-----RKKRK 682
Db 714 VKVPCDNGDCTTIGAVAAAGL---GTGAISILICIILLWSVLLFVWMMKREKRRH 770
QY 683 IKPEILLPDDTRDNVFFYEGEGGEDDYDITQLHRLGEARPEV-----VLNR 732
Db 771 TQQLLIDPEDDVRNLIKYEDEGGGEDDYDLSQLQ-----QPETDLHVLNKKVAGVAV 825
QY 733 DVAPTIPTMYRPRPA--NPDEIGNFIENLKAANTDPTAPPYDITLLVDFYEGSGSAA 790
Db 826 DERP-VGAHQYPIRPMVPHPGDIDGDFITEGLRAANDNDPTAPPYDITLLVDFYEGSGTAG 884
QY 791 SLSSLTSSASDQDDYDVLNWSGRFKKLADMYGGGEDD 829
Db 885 SVSSLSNSSSS-EDQDFDYLNDWGRFKKLADMYGGGAED 922

```

RESULT 11

```

Q7ZY8
ID Q7ZY8 PRELIMINARY; PRT; 711 AA.
AC Q7ZY8;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)

```



```

533 TLLATLIDVNDHGPVPEPRQITICNQSPVRHVLNITDKLSPTSPPQAQLTDDSDIYWT 592
540 TLVLELQVNDNAPVNERIKLCNRSAPVLISITDNDGPPAGFTVVEPQDTSKWS 599
593 AENVNEGDVTVLSLKKFLKQDYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPK 652
600 VFNETHGYSLSKPSQLEQREYNVLRVTDREGQSQAISQVCDCKGEAFQCTNKQV 659
653 GGF-----ILPVLGAVLALLFLVLLVLLVLRKKI-----KEPLLIPEDDTRDNVF 699
660 AGVPLGLGLGIGI--LLFVLLILKYLILPYLLFFSSCSCKGKTPLLSLSDVDDVRNIC 717
700 YGEGGGEEDQ 711
718 YDEECGEGDQ 729

RESULT 13
ID Q7SZW3 PRELIMINARY; PRT; 677 AA.
AC Q7SZW3;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE SI:DZ71B9.1 (Novel protein similar to zebrafish epithelial cadherin 1 (Cdh1)) (Fragment)
DE SI:DZ71B9.1.
GN Brachydanio rerio (Zebrafish) (Danio rerio).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OC NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Bates K.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL929295; CAE30436.1; -.
FT NON_TER 1
SQ SEQUENCE 677 AA; 74493 MW; AF895F3165EBA5F7 CRC64;

Query Match 28.5%; Score 1245; DB 13; Length 677;
Best Local Similarity 43.9%; Pred. No. 1.3e-78;
Matches 260; Conservative 101; Mismatches 221; Indels 10; Gaps 7;

QY 96 PP-SKRILRRHQRDWWVAPISVPENGKGFPPORLNQLKSNKDRTKIFYITGADSP 154
DB 83 PPKSSTGLKRAKRGWIIPFVSFVFNKSGKPPFMRVLVQIKSDSASQYQAYKITGEGADDP 142
QY 155 EGVAFAVEKETGMLLANKPLDREETAKEFLFGHVSNGASVEDPMNT-SIIITDQNDHKPK 214
DB 143 KGIFKIDLSLGNVNTQQLDREKASYKLVAHATGVDVNIKEPFEIVTVTDQNDKNPV 202
QY 215 FTQDTFRGSVLEGLVPGTSVMQVATDDEDDAIYVNGVAVYSIHSQPKDPKDHLMFTIHR 274
DB 203 FTQNPFNANVPEALEKGEVFTVATDADDKENTDNADISYAIISQDPPSAKP2MFAINP 262
QY 275 STGTISVLSGLDREKPEYTLTIOATDMGDSGTTTAVAVVEILDANDNAPMFDPKYE 334
DB 263 VSGGHSVLETLGDRQWERYTLVITATDMNGEGSTTGTAIVTMSNDNAPLEQTSHT 322
QY 335 AHVPENAVGHEVQRLTVDLAPNSPAWRATYLYINGMGDDGHHFTTTHPENQGIITRRK 394
DB 323 VSVPENQVGAQVANLPVTDGPESTAWSTKYRVIGSGKGFVNVSTGPSLEGVITTVK 382
QY 395 GLDFAKQHTLYVEVTNEAPVLKLPSTATIVVHVEDVNEAPVFPVPPSKVVEVQEGIP 454
DB 383 PLDYETKQYILFVIVNDKDFVGLSPYSTATVTVNVNEDVNEPPEFIPKEFISKPELDLP 442
QY 455 TCEPVCVVTAEADPKDE-NQKISYRILRDPAGLWAMPDPSGQVAVGTLDREDEQFVRNI 513
DB 443 VGSNLVIYATVDVTEKKQITRIDYDPGWLSI-TDSGQIQVKTAMDRESSN-VKDGK 500

```

```

514 YEVWLVANDNG-SPTTGTGTLTLLIDVNDHGPVPEPRQITICNQSPVRHVLNITDKOL 572
501 YKALITALLDDDDVESPATGTGTVLELQVNDNAPVINERTIKLCNRSALVLLSITDNDG 560
573 SPTSPPFQAQLTDDSDIYWTAEVNEEGDVTVLSLKKFLKQDYDVHLSLSDHGNKEQLTV 632
561 PFAAGFVSVEPQDTSKWSVFNETHGYSLSKQSLDQGEYNVLRVTDREGESQENT 620
633 IRATYCDCHGHVETCPGPKGK-----FILPVLGAVLALLFLVLLVLLVLRKK 680
621 IQASVCDCKGEAFQCTENQEAQRPFLGVLGVLGGTL-LLLLLGLLSMFLRKK 671

RESULT 14
ID Q7ZY6 PRELIMINARY; PRT; 690 AA.
AC Q7ZY6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE SI:DZ18005.2 (Novel protein similar to E-cadherin (cdh1)) (Fragment)
DE SI:DZ18005.2.
GN Brachydanio rerio (Zebrafish) (Danio rerio).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OC NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Pelan S.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL929461; CAD60789.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR Pfam; PF00028; cadherin; 5.
DR SMART; SM00112; Ca; 4.
DR PROSITE; PS00232; CADHERIN_1; 2.
DR PROSITE; PS50268; CADHERIN_2; 4.
FT NON_TER 1
SQ SEQUENCE 690 AA; 75837 MW; 70C0AB808B9F8357 CRC64;

Query Match 28.0%; Score 1221.5; DB 13; Length 690;
Best Local Similarity 39.7%; Pred. No. 6e-77;
Matches 278; Conservative 108; Mismatches 262; Indels 53; Gaps 12;

QY 48 GOALGK-VMGCGQEPALFSTNDNDFTVRNGETVQERRSLKERNPLK----- 94
DB 6 GERLGRVVESTCDGRGTIDFESTDQEIIDLNMDDGLMRRSVTLHEGLKEFSVSAWDFSGK 65
QY 95 ----IFPSKILARRK-----RDWVAPISVPENGKGFPPORLNQLKSNKDR 137
DB 66 KHTVFRVVRVDYHKDHDVNTLMTSSVQRCILFPLTLHENGKGFPPRLVQIKSDSAS 125
QY 138 DTKIFYSIITGQADSPPEGVFAVEKETGMLLANKPLDREETAKEFLFGHVSNGASVED 197
DB 126 ETPMAYKITGEGADQPKGIFQIDRLSGWVSVTQQLDREKASYKLVAHATGVDVNIKEK 185
QY 198 PMNISIIVTQDNDHKPKFTQDTFRGSVLEGLVPGTSVMQVATDDEDDAIYTVNGVAVSI 257
DB 186 PFEFIVTADQNDKPKLFTQNPFNANVPEALEKGEVFTVATDADDKENTDNADISYAI 245
QY 258 HSQEPKDPDHLMTFTIHRSTGTISVLSGLDREKPEYTLTIOATDMGDSGTTTAVVE 317
DB 246 ISQDPPSPXPNFAINPVSGG-SMLETGLDRQWERYTLVITATDMNGEGSTTGTAIVT 305
QY 318 ILDANDNAPMFPQKYEAHVBNVAVGHEVQRLTVDLAPNSPAWRATYLYINGMGDDGHHF 377
DB 306 VTDSNNNAPLFEQTLTYTVSVPNQVGAEVAKUPLFVTDGDEPESTAWSTKYRVIGGKDGFF 365

```


This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:00:03 ; Search time 70 Seconds
(without alignments)

3346.170 Million cell updates/sec

Title: US-09-916-849A-1

Perfect score: 4369

Sequence: 1 MGUPRGPLASLLQLVCWLQ.....NEWGSRFKKLADMYGGGEDD 829

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_29Jan04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4369	100.0	829	5 ABG61897	Abg61897 Prostate
2	4369	100.0	829	5 ABB81476	Abb81476 Human P-c
3	4369	100.0	829	5 ABJ05598	Abj05598 Breast ca
4	4369	100.0	829	5 AAU97492	Aau97492 Human P-c
5	4369	100.0	829	5 AAM50864	Aam50864 Cadherin
6	4369	100.0	829	6 ABP58357	Abp58357 Human P-c
7	4369	100.0	829	7 ADC15497	Adc15497 Human bas
8	4369	100.0	829	7 ADD14190	Add14190 Human src
9	4361	99.8	829	5 ABP34683	Abp34683 Metastati
10	4361	99.8	829	6 ABR58670	AbR58670 Human can
11	4361	99.8	829	6 ABUS6434	Abus6434 Lung can
12	4351	99.6	829	6 ABUS6670	Abus6670 Lung can
13	2420	55.4	882	2 AAU93375	Aay09375 Wild-type
14	2420	55.4	882	3 AAB35730	Aab35730 Human E-c
15	2420	55.4	882	4 AAB73490	Aab73490 Human E-c
16	2420	55.4	882	5 ABG96293	Abg96293 Human ova
17	2420	55.4	882	5 AAU78051	Aau78051 Human E-c
18	2388	54.7	899	5 ABB81472	Abb81472 Mouse E-c
19	2348.5	53.8	878	2 AAR85487	Aar85487 Human E-c
20	2341.5	53.6	878	2 AAR55060	Aar55060 Sequence
21	2341.5	53.6	878	5 ABB81475	Abb81475 Human E-c
22	1714	39.2	906	3 AAU70741	Aay70741 Human N-c
23	1714	39.2	906	7 ADE55478	Ades5478 Human Pro
24	1714	39.2	906	7 ADE55482	Ades5482 Human Pro
25	1714	39.2	906	7 ADE55486	Ades5486 Human Pro

26	1714	39.2	906	7 ADE55490	Ades5490 Human Pro
27	1713	39.2	906	6 ABR58643	AbR58643 Human can
28	1708	39.1	906	4 ABG14316	Abg14316 Novel hum
29	1707	39.1	906	5 ABB81474	Abb81474 Human N-c
30	1707	39.1	906	6 ABR47406	AbR47406 Breast ca
31	1704.5	39.0	912	5 ABB81471	Abb81471 Chicken N
32	1697.5	38.9	906	5 ABB57233	Abb57233 Mouse isc
33	1693	38.8	906	7 ADE55488	Ades5488 Rat Prote
34	1693	38.8	906	7 ADE55476	Ades5476 Rat Prote
35	1693	38.8	906	7 ADE55484	Ades5484 Rat Prote
36	1693	38.8	906	7 ADE55480	Ades5480 Rat Prote
37	1662	38.0	747	7 ADD14035	Add14035 Human src
38	1638	37.5	824	7 ADB64361	Adb64361 Human cad
39	1621.5	37.1	916	2 AAW25658	Aaw25658 Human cad
40	1621.5	37.1	916	2 AAW13129	Aaw13129 Full leng
41	1621.5	37.1	916	7 ADD14110	Add14110 Human src
42	1422	32.5	848	5 ABG79689	Abg79689 Tumour in
43	1321.5	30.2	814	3 AAB24089	Aab24089 Human PRO
44	1321.5	30.2	814	4 ABG30224	Abg30224 Novel hum
45	1321.5	30.2	814	5 AAE26669	Aae26669 Human cad

ALIGNMENTS

RESULT 1

ABG61897

ID ABG61897 standard; protein; 829 AA.

XX AC ABG61897;

XX DT 15-AUG-2002 (first entry)

XX DE Prostate cancer-associated protein #98.

XX KW Prostate cancer; prostate tumour tissue; human; mammal; cytostatic.

XX OS Mammalia.

XX PN WC200230268-A2.

XX PD 18-APR-2002.

XX PF 12-OCT-2001; 2001WO-US032045.

XX PR 13-OCT-2000; 2000US-00687576.

XX PR 08-DEC-2000; 2000US-00733288.

XX PR 08-DEC-2000; 2000US-00733742.

XX PR 24-JAN-2001; 2001US-02639572.

XX PR 16-MAR-2001; 2001US-0276791P.

XX PR 16-MAR-2001; 2001US-0276888P.

XX PR 06-APR-2001; 2001US-0281922P.

XX PR 24-APR-2001; 2001US-0286214P.

XX PR 30-APR-2001; 2001US-00847046.

XX PR 04-MAY-2001; 2001US-0288589P.

XX (EOSB-) EOS BIOTECHNOLOGY INC.

PA Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;

PI WPI: 2002-471335/50.

XX N-PSDB; ABK92214.

XX DR Detecting a prostate cancer-associated transcript in a cell in a patient,

XX PT useful for diagnosing prostate cancer (PC) or screening modulators of PC,

XX PT by determining if prostate cancer-associated genes are expressed in a

XX PT prostate tissue.

XX PS Claim 27; Page 384-385; 436pp; English.

XX CC The present invention relates to methods of detecting a prostate cancer-

XX CC associated transcript in a cell from a patient. The method comprises

XX CC contacting a biological sample from the patient with prostate cancer-

associated polynucleotides (designated PC genes) that selectively hybridise to a sequence that is at least 80% identical to them. The prostate cancer-associated polynucleotide sequences are differentially expressed in prostate tumour tissue or in prostate cancer and are derived from the tissues of various organisms such as humans or other mammals (e.g. mice, sheep and dogs). The methods of the invention are useful for diagnosing and treating prostate cancer in mammals. The prostate cancer-associated genes are useful for diagnosing or treating prostate cancer, as well as for identifying modulators of prostate cancer or agents that inhibit prostate cancer. The nucleic acid sequences are particularly useful in gene therapy, as a vaccine or in antisense applications.

ABG61800-ABG61944 represent prostate cancer-associated proteins.

Sequence 829 AA;

Query Match 100.0%; Score 4369; DB 5; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGLASLLLLQVCLQCAASEPCRAVFAEVTLEAGGAEQPGQALGVFMGCPG 60
 DB 1 MGLPRGLASLLLLQVCLQCAASEPCRAVFAEVTLEAGGAEQPGQALGVFMGCPG 60

QY 61 QEPALFSTDDNDFTVRGETVQERRSKERNPLKIPSKILRRKRDWVAVISPENG 120
 DB 61 QEPALFSTDDNDFTVRGETVQERRSKERNPLKIPSKILRRKRDWVAVISPENG 120

QY 121 KGPPPPQRLNQLSKNDRTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREIEAK 180
 DB 121 KGPPPPQRLNQLSKNDRTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREIEAK 180

QY 181 YELFGHAVSENGASVEDPMNISITVTQNDHKPKFTQDTRGSLGVLPGTSVMQVAT 240
 DB 181 YELFGHAVSENGASVEDPMNISITVTQNDHKPKFTQDTRGSLGVLPGTSVMQVAT 240

QY 241 DEDDAIYVNGVAYS:HSQEPKDPHDMFTIHRSTGTSIVSSGLDREKVPYTLTIOA 300
 DB 241 DEDDAIYVNGVAYS:HSQEPKDPHDMFTIHRSTGTSIVSSGLDREKVPYTLTIOA 300

QY 301 TMDGDSGTTTAVAVVILDANDNAPMDPKQYEAHVPENAVGEVQLTVDLDPNSP 360
 DB 301 TMDGDSGTTTAVAVVILDANDNAPMDPKQYEAHVPENAVGEVQLTVDLDPNSP 360

QY 361 AWRATYILMGDDGDHFTITTHPSNOGILTRKGLDPEAKNOHTLYVEVTNEAPFVLKL 420
 DB 361 AWRATYILMGDDGDHFTITTHPSNOGILTRKGLDPEAKNOHTLYVEVTNEAPFVLKL 420

QY 421 PTSTATIVHVEDNEAPFVPPSKVVEQGIPTGEPVCVYTAEPDPKXNKISYRILR 480
 DB 421 PTSTATIVHVEDNEAPFVPPSKVVEQGIPTGEPVCVYTAEPDPKXNKISYRILR 480

QY 481 DPAGLWLPDSGQVAVTGLDREDEQFVRNNIYEVWVLAMDNGSPPTTGTLTLTLD 540
 DB 481 DPAGLWLPDSGQVAVTGLDREDEQFVRNNIYEVWVLAMDNGSPPTTGTLTLTLD 540

QY 541 VNDHGVPPEPQITICNQSPVRHVLNITDKLSPHTSPFQAQLTDDSDIYTAENVNEEG 600
 DB 541 VNDHGVPPEPQITICNQSPVRHVLNITDKLSPHTSPFQAQLTDDSDIYTAENVNEEG 600

QY 601 TVVLSLKKFLKQDYDYLHSLSDHGNKEQLTVIRATVCDCHGVETCPGWKGFILPVL 660
 DB 601 TVVLSLKKFLKQDYDYLHSLSDHGNKEQLTVIRATVCDCHGVETCPGWKGFILPVL 660

QY 661 GAVLALLFLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLL 720
 DB 661 GAVLALLFLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLLVLL 720

QY 721 GLEARPEVLRNDVAPTIPTPMYRPRPANDPEIGNFIENLKAANTDTPAPPYDILLVF 780
 DB 721 GLEARPEVLRNDVAPTIPTPMYRPRPANDPEIGNFIENLKAANTDTPAPPYDILLVF 780

QY 781 DYEGSGDAASLSLTSASDQDQDYLVNWSGRFKKLADMYGGGEDD 829

DB 781 DYEGSGDAASLSLTSASDQDQDYLVNWSGRFKKLADMYGGGEDD 829

RESULT 2
 ABB81476
 ID ABB81476 standard; protein; 829 AA.
 AC ABB81476;
 XX 30-AUG-2002 (first entry)
 DT
 DE Human P-cadherin protein SEQ ID NO:50.
 XX
 KW Cadherin; alpha-catenin; cancer; beta-catenin binding domain; melanoma;
 o-catenin; colon cancer.
 OS Homo sapiens.
 XX
 PN US2002045591-A1.
 XX
 PD 18-APR-2002.
 XX
 PF 17-JUL-2001; 2001US-00905983.
 XX
 PR 26-MAY-1998; 98IL-00124650.
 PR 26-MAY-1999; 99US-00318633.
 XX
 PA (GEIG/) GEIGER B.
 PA (BENZ/) BEN-ZE'EV A.
 PA (SADO/) SADO E.
 XX
 PI Geiger B, Ben-Ze'ev A, Sadot E;
 XX
 DR WPI; 2002-499105/53.
 DR N-PSDB; ABN89393.
 XX
 PT New construct encoding soluble cytoplasmic portion of cadherin including
 beta catenin binding domain useful in treating cancer associated with
 high beta-catenin activity e.g. colon cancer and melanoma.
 XX
 PS Example 3; Page 53-55; 102pp; English.
 XX
 CC The present invention describes a pharmaceutical composition for treating
 cancer associated with abnormally high beta-catenin activity. The
 pharmaceutical composition comprises a gene therapy vehicle harbouring a
 polynucleotide that contains: (i) a nucleotide sequence encoding a
 soluble cytoplasmic portion of a cadherin which lacks a transmembrane
 portion and an extracellular portion of the cadherin, and includes a beta
 -catenin binding domain; and (b) an upstream promoter for directing
 expression of the soluble cytoplasmic portion of the cadherin in a
 mammalian cell. Also described is a pharmaceutical composition for
 treating cancer associated with abnormally high activity levels of beta-
 catenin comprising a gene therapy vehicle harbouring a polynucleotide
 that contains: (a) a nucleotide sequence encoding an o-catenin; and (b)
 an upstream promoter for directing expression of the o-catenin in a
 mammalian cell. The pharmaceutical compositions have cytoskeletal activity
 and can be used in the suppression of beta-catenin-mediated
 transactivation. They can be used for treating cancers associated with
 abnormally high activity levels of beta-catenin such as colon cancers and
 melanomas, by reducing these high activity levels of beta-catenin in
 mammalian cells. The present sequence represents human P-cadherin which
 is used in the exemplification of the present invention

Sequence 829 AA;
 Query Match 100.0%; Score 4369; DB 5; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGLASLLLLQVCLQCAASEPCRAVFAEVTLEAGGAEQPGQALGVFMGCPG 60
 DB 1 MGLPRGLASLLLLQVCLQCAASEPCRAVFAEVTLEAGGAEQPGQALGVFMGCPG 60

Qy 61 QEPALFSTNDNDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120
Db 61 QEPALFSTNDNDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120
Qy 121 KGPPFQRLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREETIAK 180
Db 121 KGPPFQRLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREETIAK 180
Qy 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTERGSLVLEGLPCTSVMOVTAT 240
Db 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTERGSLVLEGLPCTSVMOVTAT 240
Qy 241 DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTHRSTGTISVISGLDREKVPYITLTOA 300
Db 241 DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTHRSTGTISVISGLDREKVPYITLTOA 300
Qy 301 TMDGCGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGEVORLTVDLDAPNSP 360
Db 301 TMDGCGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGEVORLTVDLDAPNSP 360
Qy 361 AWRATYILMGDDGDGHFTTTHPESNQGLITTRKGLDPEAKNQHTLYVEVTNEAPFVLK 420
Db 361 AWRATYILMGDDGDGHFTTTHPESNQGLITTRKGLDPEAKNQHTLYVEVTNEAPFVLK 420
Qy 421 PTSTATIVHVEDVNEAPVFPSPKVEVQEGIPTGEPVCVYTAEDPDKENQKISYRILR 480
Db 421 PTSTATIVHVEDVNEAPVFPSPKVEVQEGIPTGEPVCVYTAEDPDKENQKISYRILR 480
Qy 481 DPAGWLAMDPSGOVTAVGTLDRDEQFVRNNIYEVNVLAMDNQSPPTTGTGTLTLLID 540
Db 481 DPAGWLAMDPSGOVTAVGTLDRDEQFVRNNIYEVNVLAMDNQSPPTTGTGTLTLLID 540
Qy 541 VNDHGPVPERQITICNQSPVRHVLNITDKDLSPTSPFOAQLTDDSDIYWTAEVNEEGD 600
Db 541 VNDHGPVPERQITICNQSPVRHVLNITDKDLSPTSPFOAQLTDDSDIYWTAEVNEEGD 600
Qy 601 TVVLSLKKFLKQDIYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL 660
Db 601 TVVLSLKKFLKQDIYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL 660
Qy 661 GAVLALLFLLLVLLVLRKKRIKEPLLLPEDDTRDNVYFYGEGGEEQDQYDITQLHR 720
Db 661 GAVLALLFLLLVLLVLRKKRIKEPLLLPEDDTRDNVYFYGEGGEEQDQYDITQLHR 720
Qy 721 GLEARPEVLNDVAPTIITPMYRPRPANDPEIGNFIENLKAANTDPTAPPYDTLLVF 780
Db 721 GLEARPEVLNDVAPTIITPMYRPRPANDPEIGNFIENLKAANTDPTAPPYDTLLVF 780
Qy 781 DYGSGSDAASLSLTSASDQDQDYDLNEMWSRFFKLDAMYGGGDD 829
Db 781 DYGSGSDAASLSLTSASDQDQDYDLNEMWSRFFKLDAMYGGGDD 829

RESULT 3
ID ABJ05598
AC ABJ05598 standard; protein; 829 AA.
XX
XX
XX
DT 14-NOV-2002 (first entry)
XX
DE Breast cancer-associated protein 63.
XX
XX Breast cancer; breast cancer-associated gene sequence; drug development;
XX pharmacogenetics; biosensor development.
XX Unidentified.
XX
XX WO200259377-A2.
XX
XX 01-AUG-2002.
XX
XX 24-JAN-2002; 2002WO-US002242.
XX
XX

XX
PR
PR
PR
PR
PR
PR
XX
PA
XX
XX
XX
XX
DR
XX
XX
PT
PT
PT
XX
PS
XX
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
XX
SQ

24-JAN-2001; 2001US-0263965P.
02-FEB-2001; 2001US-0265928P.
09-APR-2001; 2001US-00829472.
09-APR-2001; 2001US-0282698P.
04-MAY-2001; 2001US-0288590P.
29-MAY-2001; 2001US-0294443P.
(EOSB-) EOS BIOTECHNOLOGY INC.
Mack DH, Gish KC, Afar D;
WPI; 2002-583738/62.
N-PSDB; ABT07755.
Detecting a breast cancer-associated transcript in a patient's cell,
sample with a polynucleotide that selectively hybridizes with breast
cancer nucleic acids.
Disclosure; Page 405; 414pp; English.
The invention comprises a method of detecting a breast cancer-associated
transcript in a cell from a patient. The method of the invention involves
contacting a biological sample from the patient with a nucleotide that
hybridizes to one of the 69 breast cancer-associated gene sequences shown
in the specification. The method of the invention is useful in the
diagnosis or prognosis of breast cancer, and for detecting genes that are
up or down-regulated in breast cancer cells. Genes identified by the
method of the invention can be used in diagnostic purposes and also as
targets for screening for therapeutic compounds that modulate breast
cancer (e.g. hormones or antibodies). Identification of genes that are
over or under expressed in breast cancer can additionally provide high-
resolution, high-sensitivity datasets which can be used in the areas of
diagnostics, therapeutics, drug development, pharmacogenetics, protein
structure and biosensor development. Amino acid sequences ABJ05536 -
ABJ05604 represent the proteins encoded by the 69 breast cancer-
associated genes of the invention

Sequence 829 AA;

Query Match 100.0%; Score 4369; DB 5; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MGLPRGLASLLLLQVCLQCAASEPCRAVFEAEVTLGAGGAEQEPGQALCKVFMGCPG 60
Db 1 MGLPRGLASLLLLQVCLQCAASEPCRAVFEAEVTLGAGGAEQEPGQALCKVFMGCPG 60
Qy 61 QEPALFSTNDNDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120
Db 61 QEPALFSTNDNDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120
Qy 121 KGPPFQRLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREETIAK 180
Db 121 KGPPFQRLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREETIAK 180
Qy 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTERGSLVLEGLPCTSVMOVTAT 240
Db 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTERGSLVLEGLPCTSVMOVTAT 240
Qy 241 DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTHRSTGTISVISGLDREKVPYITLTOA 300
Db 241 DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTHRSTGTISVISGLDREKVPYITLTOA 300
Qy 301 TMDGCGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGEVORLTVDLDAPNSP 360
Db 301 TMDGCGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGEVORLTVDLDAPNSP 360
Qy 361 AWRATYILMGDDGDGHFTTTHPESNQGLITTRKGLDPEAKNQHTLYVEVTNEAPFVLK 420
Db 361 AWRATYILMGDDGDGHFTTTHPESNQGLITTRKGLDPEAKNQHTLYVEVTNEAPFVLK 420

QY 421 PTSTATIVVHVEDVNEAPVFPVPPSKVVEVQEGITGEPVVCVYTAEDDPKENOKISYRILR 480
 DB 421 PTSTATIVVHVEDVNEAPVFPVPPSKVVEVQEGITGEPVVCVYTAEDDPKENOKISYRILR 480
 QY 481 DPAGWLAMPDPSGGQVAVGTGLDREDEQFVRNIIYEVNVLAMDNGSPPTTGTGLLLTLID 540
 DB 481 DPAGWLAMPDPSGGQVAVGTGLDREDEQFVRNIIYEVNVLAMDNGSPPTTGTGLLLTLID 540
 QY 541 VNDHGPVPEPQIITICNOSPVRHVLNITDKLSHSTSPFOAQLTDDSDIYWTAEVNEEGD 600
 DB 541 VNDHGPVPEPQIITICNOSPVRHVLNITDKLSHSTSPFOAQLTDDSDIYWTAEVNEEGD 600
 QY 601 TVVLSLKKFLKQDITYDVHLSLSHDGNEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
 DB 601 TVVLSLKKFLKQDITYDVHLSLSHDGNEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
 QY 661 GAVLALLFLLVLLVLRKKIKIPEPLLPEDDTRDNVFFYGGEGGEEDQDITQLHR 720
 DB 661 GAVLALLFLLVLLVLRKKIKIPEPLLPEDDTRDNVFFYGGEGGEEDQDITQLHR 720
 QY 721 GLEARPEVLRNDVAPTIIPTPMYRPRPNDPEIGNFIENLKAANTDPTAPPYDITLLVF 780
 DB 721 GLEARPEVLRNDVAPTIIPTPMYRPRPNDPEIGNFIENLKAANTDPTAPPYDITLLVF 780
 QY 781 DYEGSGDAASLSLTSASDQDQDYDLNWSGRFCKLADMYGGGDD 829
 DB 781 DYEGSGDAASLSLTSASDQDQDYDLNWSGRFCKLADMYGGGDD 829

RESULT 4

AAU97492
 ID AAU97492 standard; protein; 829 AA.

AC AAU97492;

DT 13-AUG-2002 (first entry)

XX Human P-cadherin (placental cadherin) protein.

XX Human; P-cadherin; placental cadherin; solid surface; vascular graft;
 KW cell-to-cell cohesion; endothelial cell-coated surface;
 KW hydrodynamic shear; shaking container; continuous harvest system;
 KW laminar shear device.

XX Homo sapiens.

XX WO200231121-A2.

XX 18-APR-2002.

XX 15-OCT-2001; 2001WO-US032030.

XX 13-OCT-2000; 2000US-0241216P.

XX 27-OCT-2000; 2000US-0243693P.

XX 11-OCT-2001; 2001US-00975723.

XX (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.

XX Nackman G, Poty RA;

XX WPI; 2002-435443/46.

XX N-PSDB; ASK52038.

XX Populating solid surface especially biomedical devices with vascular
 PT endothelial cells by increasing cell-to-cell cohesion useful to heal
 FT vascular grafts.

XX Disclosure; Page 9-10; 38pp; English.

XX The present invention relates to a new method of populating a solid
 CC surface with cells. The method of the invention involves increasing the
 CC cell-to-cell cohesion of the cells. The method is useful for populating a
 CC solid surface with human and nonhuman cells. The invention is also useful

CC for achieving better utilisation of vascular grafts and also in any
 CC system that employs endothelial cell-coated surface susceptible to
 CC hydrodynamic shear, such as shaking container coated with endothelial
 CC cells, continuous harvest systems for harvesting the products of
 CC endothelial cells on solid surface, and laminar shear devices. The
 CC present amino acid sequence represents the human P-cadherin (placental
 CC cadherin) protein of the invention

XX Sequence 829 AA;

Query Match 100.0%; Score 4369; DB 5; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASLLLLQVCWLQCAASEPCRAVREAEVTLQAGAEQPCQALGKVFMCQPG 60
 DB 1 MGLPRGPLASLLLLQVCWLQCAASEPCRAVREAEVTLQAGAEQPCQALGKVFMCQPG 60
 QY 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDVWVAPISVPENG 120
 DB 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDVWVAPISVPENG 120
 QY 121 KQFFQRLNQLKSNKDRDTKIFYSITGSGADSPPEGVFAVEKETGWLNLKPLDREIAK 180
 DB 121 KQFFQRLNQLKSNKDRDTKIFYSITGSGADSPPEGVFAVEKETGWLNLKPLDREIAK 180
 QY 181 YELFGHVSSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSLVLEGVLPCTSYMQVTAT 240
 DB 181 YELFGHVSSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSLVLEGVLPCTSYMQVTAT 240
 QY 241 DEDDAIYNGVWAYSIHSQEPKOPHDLMTFHSTGTISVISSGLDREKPEYTLTIOA 300
 DB 241 DEDDAIYNGVWAYSIHSQEPKOPHDLMTFHSTGTISVISSGLDREKPEYTLTIOA 300
 QY 301 TDMGDGSGTTTAVAVVEILDANDNAPMFDPOKYBAHVPENAVGHVEQSLTVDLDAENSP 360
 DB 301 TDMGDGSGTTTAVAVVEILDANDNAPMFDPOKYBAHVPENAVGHVEQSLTVDLDAENSP 360
 QY 361 AWRATYLMGDDGDHFTTTHPESNQILITRGLDPEAKNQHTLYVEVTNEAPFVLKL 420
 DB 361 AWRATYLMGDDGDHFTTTHPESNQILITRGLDPEAKNQHTLYVEVTNEAPFVLKL 420
 QY 421 PTSTATIVVHVEDVNEAPVFPVPPSKVVEVQEGITGEPVVCVYTAEDDPKENOKISYRILR 480
 DB 421 PTSTATIVVHVEDVNEAPVFPVPPSKVVEVQEGITGEPVVCVYTAEDDPKENOKISYRILR 480
 QY 481 DPAGWLAMPDPSGGQVAVGTGLDREDEQFVRNIIYEVNVLAMDNGSPPTTGTGLLLTLID 540
 DB 481 DPAGWLAMPDPSGGQVAVGTGLDREDEQFVRNIIYEVNVLAMDNGSPPTTGTGLLLTLID 540
 QY 541 VNDHGPVPEPQIITICNOSPVRHVLNITDKLSHSTSPFOAQLTDDSDIYWTAEVNEEGD 600
 DB 541 VNDHGPVPEPQIITICNOSPVRHVLNITDKLSHSTSPFOAQLTDDSDIYWTAEVNEEGD 600
 QY 601 TVVLSLKKFLKQDITYDVHLSLSHDGNEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
 DB 601 TVVLSLKKFLKQDITYDVHLSLSHDGNEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
 QY 661 GAVLALLFLLVLLVLRKKIKIPEPLLPEDDTRDNVFFYGGEGGEEDQDITQLHR 720
 DB 661 GAVLALLFLLVLLVLRKKIKIPEPLLPEDDTRDNVFFYGGEGGEEDQDITQLHR 720
 QY 721 GLEARPEVLRNDVAPTIIPTPMYRPRPNDPEIGNFIENLKAANTDPTAPPYDITLLVF 780
 DB 721 GLEARPEVLRNDVAPTIIPTPMYRPRPNDPEIGNFIENLKAANTDPTAPPYDITLLVF 780
 QY 781 DYEGSGDAASLSLTSASDQDQDYDLNWSGRFCKLADMYGGGDD 829
 DB 781 DYEGSGDAASLSLTSASDQDQDYDLNWSGRFCKLADMYGGGDD 829

RESULT 5
 AAM50864

ID	AM50864 standard; protein; 829 AA.	Db	61	QEPALFSTDNDDFTVRNGETVQERRSLKERPLKIFPSKRILRRHKRWVAVIPSPENG	120
XX	AM50864;				
AC					
XX	07-MAY-2002 (first entry)	Qy	121	KGPPFQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLILNKPDLREETAK	180
DT					
XX		Db	121	KGPPFQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLILNKPDLREETAK	180
XX					
DE	Cadherin 3, basal cell marker.	Qy	181	YELFGHAYSENGASVEDPWNISIIVTQNDKHKPKFTQDTRFGSVLEGLVPGTSMQVAT	240
KW	Cadherin 3; P-cadherin; human; basal cell; marker; breast cancer; tumour;				
KW	diagnosis; prognosis; classification; gene therapy.	Db	181	YELFGHAYSENGASVEDPWNISIIVTQNDKHKPKFTQDTRFGSVLEGLVPGTSMQVAT	240
XX					
OS	Homo sapiens.	Qy	241	DEDDAIYTYNGVWVAYSIHQBPKDPHDLMTIHRSTGTISVISGLDREKVEPYTLTIOA	300
XX					
XX	WO200208765-A2.	Db	241	DEDDAIYTYNGVWVAYSIHQBPKDPHDLMTIHRSTGTISVISGLDREKVEPYTLTIOA	300
XX					
PD	31-JAN-2002.	Qy	301	TDMDGGSGTTTAVAVVEILDANDNAPMEDPKQYEAHVPENAVGHVEQBLTVDLDAFNSP	360
XX					
XX	26-JUL-2001; 2001WO-US023843.	Db	301	TDMDGGSGTTTAVAVVEILDANDNAPMEDPKQYEAHVPENAVGHVEQBLTVDLDAFNSP	360
XX					
PR	26-JUL-2000; 2000US-0220967P.	Qy	361	AMRATYLMGSGDGDHFTITTHPESNQGLITTRKGLDPEAKNQHTLYVEVINEAPFVLKL	420
XX					
XX	(STRD) UNIV STANFORD.	Db	361	AMRATYLMGSGDGDHFTITTHPESNQGLITTRKGLDPEAKNQHTLYVEVINEAPFVLKL	420
PA	(GENO-) APPLIED GENOMICS INC.				
PA		Qy	421	PTSTATIVVHVEDVNEAPVFPVPPSKVVEVQGGIPTGEPVCVYTABDPPKENQKISYRILR	480
PI	Botstein D, Brown PO, Perou C, Ross D, Van De Rijn M, Ring B;				
PI	Seitz R;	Db	421	PTSTATIVVHVEDVNEAPVFPVPPSKVVEVQGGIPTGEPVCVYTABDPPKENQKISYRILR	480
XX					
DR	WPI; 2002-206094/26.	Qy	481	DPAGLWMDPDSGQVTAVTGLDREDEQFVRNIIYVMVLAWDNGSPPTTGTITLLTILID	540
XX					
XX	Classifying and treating breast tumors associated with the expression and	Db	481	DPAGLWMDPDSGQVTAVTGLDREDEQFVRNIIYVMVLAWDNGSPPTTGTITLLTILID	540
PT	activity of cadherin 3 or P-cadherin, matrix metalloproteinase 14 and/or				
PT	cadherin EGF LAG seven-pass G-type receptor 2 or EGF-like Domain,	Qy	541	VNDHGFVPEPRQITICNQSPVRHVLNITDKLSPTSPFQALTDSDIYWTAEVNEEGD	600
PT	Multiple 2.	Db	541	VNDHGFVPEPRQITICNQSPVRHVLNITDKLSPTSPFQALTDSDIYWTAEVNEEGD	600
XX					
PS	Claim 1; Fig 1A; 741pp; English.	Qy	601	TVVLSIKKFLKODTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGGFILLPVL	660
XX					
CC	The present sequence is that of human cadherin 3 (or P-cadherin). The	Db	601	TVVLSIKKFLKODTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGGFILLPVL	660
CC	invention provides new reagents and methods for the management (e.g.				
CC	detection, classification, provision of diagnostic and prognostic	Qy	661	GAVLALLFLLVLLVLLVRKKKIKKEPFLLLPEDDTRDNVYFYEGEGGSEDDQDITQLHR	720
CC	information, treatment, etc.) of breast cancer. cDNA microarray	Db	661	GAVLALLFLLVLLVLLVRKKKIKKEPFLLLPEDDTRDNVYFYEGEGGSEDDQDITQLHR	720
CC	technology was used to identify genes whose expression profile across a				
CC	large group of tumour samples correlated with that of cytokeratin 5 and	Qy	721	GLEARPEVLRNDVAPTIIPTMYRPRPANDEIGNFIENLKAANTDPTAPPDYTLVLF	780
CC	cytokeratin 17, markers for basal cells of the normal mammary lactation	Db	721	GLEARPEVLRNDVAPTIIPTMYRPRPANDEIGNFIENLKAANTDPTAPPDYTLVLF	780
CC	gland. Tumours that express cytokeratin 5/6 and/or 17 have a poor				
CC	prognosis relative to tumours overall. The basal marker genes identified	Qy	781	DYEGSGSDAASLSLTSSASDQDQDYDYLNEWGSRFKKLADMYGGGDD	829
CC	included those encoding cadherin 3, matrix metalloproteinase 14 (see	Db	781	DYEGSGSDAASLSLTSSASDQDQDYDYLNEWGSRFKKLADMYGGGDD	829
CC	AM50865), and cadherin EGF LAG seven-pass G-type receptor 2 (see				
CC	AM50866). Detection of these markers allows classification of a tumour	Qy			
CC	sample as belonging to a basal tumour subclass. By providing reagents	Db			
CC	that may reliably be used to classify tumours as belonging to a basal				
CC	subclass, the invention enables a variety of methods for improving				
CC	therapeutic options for patients with breast cancer and for				
CC	individualising therapy. Potential therapeutic agents include compounds				
CC	that modulate breast cell marker genes or that modulate				
CC	polypeptides encoded by these genes. In particular, therapeutic agents				
CC	include antibodies that specifically bind to the basal marker proteins				
CC	and which recognise basal cells of normal mammary lactation glands. The				
CC	antibodies can be used either by themselves or when conjugated to, or				
CC	delivered with, another molecule such as a toxic compound. The likelihood				
CC	that cadherin 3 is membrane-bound makes it an attractive candidate for				
CC	antibody therapeutics				
XX					
SQ	Sequence 829 AA;				
	Query Match				
	Best Local Similarity 100.0%; Score 4369; DB 5; Length 829;				
	Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	1 MGLPRGPLASILLQVCWLCQAASEPCRAVFREAVTLEAGGAEQBPQALGVFMGCPG				60
Db	1 MGLPRGPLASILLQVCWLCQAASEPCRAVFREAVTLEAGGAEQBPQALGVFMGCPG				60
Qy	61 QEPALFSTDNDDFTVRNGETVQERRSLKERPLKIFPSKRILRRHKRWVAVIPSPENG				120

PR 31-MAY-2001; 2001US-0294225P.
 XX (CHIR) CHIRON CORP.
 PA Reinhard C, Klinger J, Jefferson AB, Escobedo J, Randazo F;
 PI Winter J, Goodson R;
 XX N-PSDB; ABZ24736.
 DR WPI: 2003-140501/13.
 XX Inhibiting migration and proliferation of P-cadherin expressing cancer.
 PT for treating cancer, especially digestive cancer, characterized by
 PT overexpression of P-cadherin, involves administering a P-cadherin
 PT antagonist.
 XX
 PS Disclosure; Page 10-11; 129pp; English.
 XX
 CC The present sequence is the protein sequence for human P-cadherin
 CC (placental cadherin), a calcium-dependent cellular adhesion protein. The
 CC invention provides methods of treating or diagnosing cancers involving P-
 CC cadherin expression using ligands that target P-cadherin, especially
 CC human anti-P-cadherin antibodies. A claimed method of treating a cancer
 CC characterised by the overexpression and/or upregulation of P-cadherin
 CC comprises the administration of a P-cadherin antagonist, optionally
 CC conjugated to a therapeutic agent. The migration, adhesion and/or
 CC proliferation of the cancer is inhibited, and the method is especially
 CC useful for treating or preventing a digestive cancer such as colon or
 CC colorectal cancer. The antagonists may be an anti-P-cadherin antibody or
 CC its fragment, a ribozyme or antisense oligonucleotide. A transgenic
 CC animal that expresses a recombinant (human) antibody that specifically
 CC binds P-cadherin is also claimed. A method of determining the presence of
 CC cancer involves determining the expression level of P-cadherin in a cell
 CC sample. Also provided are screens for identifying anti-P-cadherin
 CC antibodies have therapeutic activity
 XX
 SQ Sequence 829 AA;
 Query Match 100.0%; Score 4369; DB 6; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGLPRGPLASLLLLQVLCWCAASEPCRAVFAEVTLEAGGAEQEPGQALGVFMGCPG 60
 DB 1 MGLPRGPLASLLLLQVLCWCAASEPCRAVFAEVTLEAGGAEQEPGQALGVFMGCPG 60
 QY 61 QEPALFSTDDNDDFTVRNGETVQERRSLKERNPLKIPPSKILRRHKRDWVAVIPSPENG 120
 DB 61 QEPALFSTDDNDDFTVRNGETVQERRSLKERNPLKIPPSKILRRHKRDWVAVIPSPENG 120
 QY 121 KGPFPQRLNQLKSNKRDRTKIFYSITGPGADSPPEGFAVEKETGMLLNKPLDRREIAK 180
 DB 121 KGPFPQRLNQLKSNKRDRTKIFYSITGPGADSPPEGFAVEKETGMLLNKPLDRREIAK 180
 QY 181 YELFGHAVSENGASVEDPMNISIITVDQNDHKPKFTQDTPRGSLVGLPCTSMQVAT 240
 DB 181 YELFGHAVSENGASVEDPMNISIITVDQNDHKPKFTQDTPRGSLVGLPCTSMQVAT 240
 QY 241 DEDDAIYTYNGVAVYSIHSQEPKDPHOLMETIHSTGTISVISGLDREKVPYTLTQIA 300
 DB 241 DEDDAIYTYNGVAVYSIHSQEPKDPHOLMETIHSTGTISVISGLDREKVPYTLTQIA 300
 QY 301 TDMGDSGTTTAVAVVEILDANDNAPFDPQKYEAHVPENAVGEVQRLTVDLDAVNSP 360
 DB 301 TDMGDSGTTTAVAVVEILDANDNAPFDPQKYEAHVPENAVGEVQRLTVDLDAVNSP 360
 QY 361 AWRTATYLMGDDGDDHTTITHPSNQGLITRKLDFEAKQHTLYVETNEAPFVKL 420
 DB 361 AWRTATYLMGDDGDDHTTITHPSNQGLITRKLDFEAKQHTLYVETNEAPFVKL 420
 QY 421 PTSTATIVVHVEDVNEAPFVPPSKVVEVQEGIPTGFPVCVYTAEDPDKENQISYRILR 480
 DB 421 PTSTATIVVHVEDVNEAPFVPPSKVVEVQEGIPTGFPVCVYTAEDPDKENQISYRILR 480

QY 481 DPAGWLAMPDPSGGQVAVGTLDREDEQFVRNNIYEVWVYAMDNQSPPTTGTGTLTLLID 540
 DB 481 DPAGWLAMPDPSGGQVAVGTLDREDEQFVRNNIYEVWVYAMDNQSPPTTGTGTLTLLID 540
 QY 541 VNDHGVPPEPRQITICNOSPVRHVLNITDKDLSHTSPFQALTTDDSDIYWTAEVNEGD 600
 DB 541 VNDHGVPPEPRQITICNOSPVRHVLNITDKDLSHTSPFQALTTDDSDIYWTAEVNEGD 600
 QY 601 TVVLSLKXFLKQDITYVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILLPV 660
 DB 601 TVVLSLKXFLKQDITYVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILLPV 660
 QY 661 GAVLALLFLLVLLVLLVLRKKIKIPEPLLPEDDTRDNVFFYGGEGGEDDQDITQIHR 720
 DB 661 GAVLALLFLLVLLVLLVLRKKIKIPEPLLPEDDTRDNVFFYGGEGGEDDQDITQIHR 720
 QY 721 GLEAPPEVVLNDVAPTIIPTPMYRPRANDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 DB 721 GLEAPPEVVLNDVAPTIIPTPMYRPRANDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 QY 781 DYEGSGSDAASLSLTSSASDQDQDYDYLNEWGSRFKKLADMYGGGDD 829
 DB 781 DYEGSGSDAASLSLTSSASDQDQDYDYLNEWGSRFKKLADMYGGGDD 829
 RESULT 7
 ADC15497
 ID ADC15497 standard; protein; 829 AA.
 XX AC ADC15497;
 XX AC ADC15497;
 DT 18-DEC-2003 (first entry)
 XX Human basal cell marker Cadherin-3.
 DE Human; basal cell marker; cadherin-3; tumour; P-cadherin;
 KW matrix metalloproteinase 14;
 KW cadherin EGF LAG seven-pass G-type receptor 2;
 KW EGF-like Domain Multiple 2; mammary lactation gland; breast tumour;
 KW cancer; cytostatic.
 XX Homo sapiens.
 OS US2003086934-A1.
 EN 08-MAY-2003.
 PD 26-JUL-2001; 2001US-00916849.
 PF 26-JUL-2000; 2000US-0220967P.
 PR (BOTS/) BOTSTEIN D.
 PA (BROW/) BROWN P O.
 PA (PERO/) PEROU C M.
 PA (RING/) RING B.
 PA (ROSS/) ROSS D.
 PA (SEIT/) SEITZ R.
 PA (VRIJ/) VAN DE RIJN J M.
 XX Botstein D, Brown PO, Perou CM, Ring B, Ross D, Seitz R;
 PI Van De Riijn JM;
 XX WPI: 2003-786931/74.
 XX Classifying breast tumor by detecting expression or activity of gene
 PT encoding cadherin 3, matrix metalloproteinase 14, cadherin EGF LAG seven-
 PT pass G-type receptor 2 in tumor sample.
 XX Claim 1; SEQ ID NO 1; 89pp; English.
 XX The invention relates to classifying (M1) a tumour comprising detecting
 CC expression or activity of a gene encoding cadherin 3 or P-cadherin
 CC , matrix metalloproteinase 14 , or cadherin EGF LAG seven-pass G-type

receptor 2 or EGF-like domain, Multiple 2 (P1-3) or at least two genes chosen from gene encoding (P1), (P2) or (P3) in the tumour sample, and classifying tumour as belonging to a tumour subclass based on results of detecting step. Also included are testing, or predictive information based P1-3 and providing diagnostic, prognostic, or predictive information based on the detecting step, stratifying the subject for a clinical trial based on the detecting step, or selecting a treatment based on the detecting step, an antibody that specifically binds to an epitope found in (P1), (P2) or (P3) (where the antibody recognises basal cells in normal mammary lactation glands), methods of testing a compound or a combination of compounds for activity against tumours. The method (M1) is useful for classifying a tumour (breast tumour), as belonging to a tumour subclass such as basal tumour subclass. The antibody is useful for treating a subject in need of treatment for cancer. The tumour is a breast tumour, and where the method further involves identifying the tumour as belonging to a basal tumour subclass. The antibody is conjugated with a toxic molecule. The present sequence is human Cadherin-3.

Sequence 829 AA;

Query Match 100.0%; Score 4369; DB 7; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGLASILLQVCWLCQAASEPCRAVFREAEVLEAGAEQEPGQALGVFMGCPG 60
 DB 1 MGLPRGLASILLQVCWLCQAASEPCRAVFREAEVLEAGAEQEPGQALGVFMGCPG 60

QY 61 QEPALFSTNDNDFTVRNGETVQERRSLKERNPLKIPPSKILRRHKRDWVAIPSPENG 120
 DB 61 QEPALFSTNDNDFTVRNGETVQERRSLKERNPLKIPPSKILRRHKRDWVAIPSPENG 120

QY 121 KGFPFQRLNQLKSKNDRDKIFYSITGPGADSPGEGVFAVEKETGMLLNKPLDREIAK 180
 DB 121 KGFPFQRLNQLKSKNDRDKIFYSITGPGADSPGEGVFAVEKETGMLLNKPLDREIAK 180

QY 181 YELFGHAVSENGASVEDPMNISIIITDQNDHKPKFTQDTFRGSLVLEGVLPSTVMQVAT 240
 DB 181 YELFGHAVSENGASVEDPMNISIIITDQNDHKPKFTQDTFRGSLVLEGVLPSTVMQVAT 240

QY 241 DEDDAIYTVNGVAYSHSQEPKDPHLMFTIHRSTGTSIVSSGLDREKVPYITLTIQA 300
 DB 241 DEDDAIYTVNGVAYSHSQEPKDPHLMFTIHRSTGTSIVSSGLDREKVPYITLTIQA 300

QY 301 TMDGSGSTTAVAVVEILDANDNAPFDPQKYEAHVPENAVGVHVEQLTVTDLDPNSP 360
 DB 301 TMDGSGSTTAVAVVEILDANDNAPFDPQKYEAHVPENAVGVHVEQLTVTDLDPNSP 360

QY 361 AWRATYILINGDDGDHFTIITHPESNGOILTRKGLDFAKNQHTLYVEVTNEAPVFKL 420
 DB 361 AWRATYILINGDDGDHFTIITHPESNGOILTRKGLDFAKNQHTLYVEVTNEAPVFKL 420

QY 421 PTSTATVHVHVEDVNEAPVFPVPSKVVEOEGIPTGEPVCVYTAEDPDKENQKISYRIIR 480
 DB 421 PTSTATVHVHVEDVNEAPVFPVPSKVVEOEGIPTGEPVCVYTAEDPDKENQKISYRIIR 480

QY 481 DPAGWLAMPDQSGVTAAGTLDREDEQFVRNNIYEVVWVLANMGSPPTTGTGLTLTLD 540
 DB 481 DPAGWLAMPDQSGVTAAGTLDREDEQFVRNNIYEVVWVLANMGSPPTTGTGLTLTLD 540

QY 541 VNDHGPVPEPQITICNSQSPRVHVNITDKLSPTSPFOAQLTDSDIYVTAEVNEEGD 600
 DB 541 VNDHGPVPEPQITICNSQSPRVHVNITDKLSPTSPFOAQLTDSDIYVTAEVNEEGD 600

QY 601 TVVLSLKKFKLQDITVDVHLSLSHGHKNEQLTVIRATVCDCHGVETCPGPKGFFILPVL 660
 DB 601 TVVLSLKKFKLQDITVDVHLSLSHGHKNEQLTVIRATVCDCHGVETCPGPKGFFILPVL 660

QY 661 GAVLALLFLLVLLVLLVLRKKRKIKPELLLPEDDTRDNVFFYBEGGEGEDQDYITQLHR 720
 DB 661 GAVLALLFLLVLLVLLVLRKKRKIKPELLLPEDDTRDNVFFYBEGGEGEDQDYITQLHR 720

QY 721 GLEARPEVVLNRNDVAPTIIPTMYRPRANPDEIGNFIENLKAANTDPTAPPYDTLLVF 780

DB 721 GLEARPEVVLNRNDVAPTIIPTMYRPRANPDEIGNFIENLKAANTDPTAPPYDTLLVF 780

QY 781 DYEGSGSDAASISSTSSASDQDQDYDYLNEGWSRPFKKLADMYGGEDD 829

DB 781 DYEGSGSDAASISSTSSASDQDQDYDYLNEGWSRPFKKLADMYGGEDD 829

RESULT 8
 ADD14190
 ID ADD14190 standard; protein; 829 AA.
 AC ADD14190;
 XX 01-JAN-2004 (first entry)
 DE Human src biomarker polypeptide SEQ ID NO:379.
 XX predictor set; protein tyrosine kinase activity modulator;
 KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;
 KW Gene therapy; drug sensitivity; genetic profile; cancer; human.
 XX Homo sapiens.
 OS WO2003062395-A2.
 PN 31-JUL-2003.
 PD 17-JAN-2003; 2003WO-US001981.
 PF 18-JAN-2002; 2002US-0350061P.
 PR (BRIM) BRISTOL-MYERS SQUIBB CO.
 PA Huang F, Fairchild CR, Lee FY, Shaw P;
 PI WPI; 2003-636735/60.
 DR N-PSDB; ADD14795.
 XX New polynucleotides and polypeptides for predicting the activity of compounds that interact with protein tyrosine kinases and/or protein tyrosine kinase pathways.
 PS Claim 10; SEQ ID NO 379; 139pp; English.
 CC The present invention describes a predictor set comprising a plurality of polynucleotides or polypeptides whose expression pattern is predictive of the response of cells to treatment with a compound that modulates protein tyrosine kinase activity or members of the protein tyrosine kinase pathway. Also described: (1) predicting whether a compound is capable of modulating the activity of cells, comprising obtaining a sample of cells, determining whether the cells express a plurality of markers, and correlating the expression of the markers to the compound's ability to modulate the activity of the cells; (2) a plurality of cell lines for identifying polynucleotides and polypeptides whose expression levels correlate with compound sensitivity or resistance of cells associated with a disease state; and (3) identifying polynucleotides and polypeptides that predict compound sensitivity or resistance of cells associated with a disease state, comprising subjecting the plurality of cell lines to one or more compounds, analysing the expression pattern of a microarray of polynucleotides or polypeptides, and selecting polynucleotides or polypeptides that predict the sensitivity or resistance of cells associated with a disease state by using the expression pattern of the microarray. The polynucleotides and polypeptides have cytostatic activities, and can be used in gene therapy. The polynucleotides and polypeptides are useful in predicting the activity of compounds that interact with protein tyrosine kinases and/or protein tyrosine kinase pathways. These may be used in determining drug sensitivity in patients to allow the development of individualized genetic profiles which aid in treating diseases and disorders (e.g. cancer) based on patient response at a molecular level. The present sequence is used in the exemplification of the present invention.

SQ		Sequence 829 AA;	
Query Match		100.0%; Score 4369; DB 7; Length 829;	
Best Local Similarity		100.0%; Pred. No. 0;	
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	MGLPRGPLASLLLLQVLCWQCAASEPCRAVPREAEVTLEAGAGAEQEPQALGKVFMC	60
Db	1	MGLPRGPLASLLLLQVLCWQCAASEPCRAVPREAEVTLEAGAGAEQEPQALGKVFMC	60
Qy	61	QEPALFSTDNDDFTVRNGETVQERSLKERNPLKIPSKRILRRHKRDWVAPISVPENG	120
Db	61	QEPALFSTDNDDFTVRNGETVQERSLKERNPLKIPSKRILRRHKRDWVAPISVPENG	120
Qy	121	KGPFQRLNQLKSNKDRDTKIFYSTGPGADSPPEGVFAVEKETGMLLNKPLDREETAK	180
Db	121	KGPFQRLNQLKSNKDRDTKIFYSTGPGADSPPEGVFAVEKETGMLLNKPLDREETAK	180
Qy	181	YELFGHVSNGASVEDPMNISIIITDQNDHKPKFTQDTPRGVLEGLVPGTSVMQVAT	240
Db	181	YELFGHVSNGASVEDPMNISIIITDQNDHKPKFTQDTPRGVLEGLVPGTSVMQVAT	240
Qy	241	DEDDAIYTYNGVVAISYHSQEPKDPHDLMTIHRSTGTISVSSGLDREKVPYTLTIOA	300
Db	241	DEDDAIYTYNGVVAISYHSQEPKDPHDLMTIHRSTGTISVSSGLDREKVPYTLTIOA	300
Qy	301	TMGDGSGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHVQRLTVDLDA	360
Db	301	TMGDGSGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHVQRLTVDLDA	360
Qy	361	AWRATYLLMGDDGDHFTITTHPESNQILITTRKGLDFEAKNQHTLYVEVTNEAPFVLKL	420
Db	361	AWRATYLLMGDDGDHFTITTHPESNQILITTRKGLDFEAKNQHTLYVEVTNEAPFVLKL	420
KW		Colorectal cancer; metastasis; differential expression; cytostatic; diagnosis; gene therapy; vaccine.	
XX		Homo sapiens.	
OS		WO200268677-A2.	
XX		06-SEP-2002.	
PF		27-FEB-2002; 2002WO-US006001.	
XX		27-FEB-2001; 2001US-0272206P.	
PR		02-APR-2001; 2001US-0281149P.	
PR		17-APR-2001; 2001US-0284555P.	
XX		(EOSB-) EOS BIOTECHNOLOGY INC.	
PA		(UYCA-) UNIV CASE WESTERN RESERVE.	
XX		Mack DH, Markowitz SD;	
PI		WPI: 2002-698677/75.	
XX		N-PSDB; ABQ81547.	
DR		New genes that are up- or down-regulated in colorectal cancer, useful for diagnosing colorectal cancer in a subject, or for identifying modulators of colorectal cancer-associated proteins and genes for treating colorectal cancer.	
PT		Claim 8; Page 245; 260pp; English.	
XX		The present sequence is the protein sequence of a human polypeptide encoded by a gene that exhibits increased expression in colon cancer-derived metastases compared to normal colon tissue. It is an example of claimed polypeptides that are encoded by genes which are differentially expressed in metastatic colorectal cancer cells. Such polypeptides are useful in diagnostic and prognostic assays, for raising antibodies useful e.g. in immunotherapy, and in screening for modulator compounds of therapeutic value	
CC		Query Match 99.8%; Score 4361; DB 5; Length 829;	
XX		Best Local Similarity 99.9%; Pred. No. 0;	
SQ		Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
Qy	1	MGLPRGPLASLLLLQVLCWQCAASEPCRAVPREAEVTLEAGAGAEQEPQALGKVFMC	60
Db	1	MGLPRGPLASLLLLQVLCWQCAASEPCRAVPREAEVTLEAGAGAEQEPQALGKVFMC	60
Qy	61	QEPALFSTDNDDFTVRNGETVQERSLKERNPLKIPSKRILRRHKRDWVAPISVPENG	120
Db	61	QEPALFSTDNDDFTVRNGETVQERSLKERNPLKIPSKRILRRHKRDWVAPISVPENG	120
Qy	121	KGPFQRLNQLKSNKDRDTKIFYSTGPGADSPPEGVFAVEKETGMLLNKPLDREETAK	180
Db	121	KGPFQRLNQLKSNKDRDTKIFYSTGPGADSPPEGVFAVEKETGMLLNKPLDREETAK	180
Qy	181	YELFGHVSNGASVEDPMNISIIITDQNDHKPKFTQDTPRGVLEGLVPGTSVMQVAT	240
Db	181	YELFGHVSNGASVEDPMNISIIITDQNDHKPKFTQDTPRGVLEGLVPGTSVMQVAT	240
Qy	241	DEDDAIYTYNGVVAISYHSQEPKDPHDLMTIHRSTGTISVSSGLDREKVPYTLTIOA	300
Db	241	DEDDAIYTYNGVVAISYHSQEPKDPHDLMTIHRSTGTISVSSGLDREKVPYTLTIOA	300
Qy	301	TMGDGSGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHVQRLTVDLDA	360
Db	301	TMGDGSGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHVQRLTVDLDA	360
Qy	361	AWRATYLLMGDDGDHFTITTHPESNQILITTRKGLDFEAKNQHTLYVEVTNEAPFVLKL	420
Db	361	AWRATYLLMGDDGDHFTITTHPESNQILITTRKGLDFEAKNQHTLYVEVTNEAPFVLKL	420
Qy	421	PTSTATIVVHVEDVNEAPFVPPSKVVEQEGPTGEPVCVYTAEDPDKENKISYRILR	480
Db	421	PTSTATIVVHVEDVNEAPFVPPSKVVEQEGPTGEPVCVYTAEDPDKENKISYRILR	480
Qy	481	DPAGWLAMDPSGGVTVAGTLDRDEQFVRNNIYEVNVLAMDNQSPPTTGTGILLTLID	540
Db	481	DPAGWLAMDPSGGVTVAGTLDRDEQFVRNNIYEVNVLAMDNQSPPTTGTGILLTLID	540
Qy	541	VNDHGPVPEPQITICNQSPVRHVLNITDKDLSHTSPFOAQLTDDSDIYWTAEVNEEGD	600
Db	541	VNDHGPVPEPQITICNQSPVRHVLNITDKDLSHTSPFOAQLTDDSDIYWTAEVNEEGD	600
Qy	601	TVLTLKFLKQDQYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL	660
Db	601	TVLTLKFLKQDQYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL	660
Qy	661	CAVLALLFLLLVLLVLLKKEKIKEPLLPEDDTRDNVYFYGEGGEEQDQYDITQLHR	720
Db	661	CAVLALLFLLLVLLVLLKKEKIKEPLLPEDDTRDNVYFYGEGGEEQDQYDITQLHR	720
Qy	721	GLEARPEVLRNDVAPITPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPVDTLLVF	780
Db	721	GLEARPEVLRNDVAPITPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPVDTLLVF	780
Qy	781	DYEGSGDAASLSLTSSASQDQDYDYLNEWGSFRFKKLADMYGGGDD 829	
Db	781	DYEGSGDAASLSLTSSASQDQDYDYLNEWGSFRFKKLADMYGGGDD 829	
RESULT 9		ABP54683	
ID		ABP54683 standard; protein; 829 AA.	
XX		AC ABP54683;	
XX		30-DEC-2002 (first entry)	
DE		Metastatic colorectal cancer-associated polypeptide.	
XX			

QY	421	PTSTATIVHVEDVNEAPVFPVPSKVVEVQEGIPTEGPEVCYVYTAEDDPKENQKISYRLR	480
DB	421	PTSTATIVHVEDVNEAPVFPVPSKVVEVQEGIPTEGPEVCYVYTAEDDPKENQKISYRLR	480
QY	481	DPAGWLAMPDSCQVTAAGTLDREDEQFVRNNIYEVWLAMDNGSPPTTGTGTLILLTLD	540
DB	481	DPAGWLAMPDSCQVTAAGTLDREDEQFVRNNIYEVWLAMDNGSPPTTGTGTLILLTLD	540
QY	541	VNDHGVPVPRQITTCNQSPVRHVLIITDKOLSPHTSPFQAQLTDDSDIYWTAEVNEEGD	600
DB	541	VNDHGVPVPRQITTCNQSPVRHVLIITDKOLSPHTSPFQAQLTDDSDIYWTAEVNEEGD	600
QY	601	TVVLSLKKELKQDVTYDVHLSLDHGNGKQLTVIRATVCDCHGHVETCPGPKWGGFILLPVL	660
DB	601	TVVLSLKKELKQDVTYDVHLSLDHGNGKQLTVIRATVCDCHGHVETCPGPKWGGFILLPVL	660
QY	661	GAVLALLFLLVLLVLRKKRRIKEPILLPDDTRDNVYFYGEGGGEEDQDYDITQLHR	720
DB	661	GAVLALLFLLVLLVLRKKRRIKEPILLPDDTRDNVYFYGEGGGEEDQDYDITQLHR	720
QY	721	GLEAREPVLRNDVAPTIIPTPMYPRPANDEIGNFIIENLKAANTDPTAPPYDILLVF	780
DB	721	GLEAREPVLRNDVAPTIIPTPMYPRPANDEIGNFIIENLKAANTDPTAPPYDILLVF	780
QY	781	DYEGSGSDAASLSLTSSTASDQDQDYDLNENWGSRRFKKLADMYGGGDD	829
DB	781	DYEGSGSDAASLSLTSSTASDQDQDYDLNENWGSRRFKKLADMYGGGDD	829
RESULT 10			
ABR58670			
ABR58670 standard; protein; 829 AA.			
XX	AC	ABR58670;	
XX	DT	09-JUL-2003 (first entry)	
XX	DE	Human cancer related protein SEQ ID NO:327.	
XX	KW	Human; cancer; diagnosis; screening; modulator; leukaemia; ischaemia;	
XX	KW	heart disease; atherosclerosis; endometriosis.	
XX	OS	Homo sapiens.	
XX	PN	WO2003025138-A2.	
XX	XX	27-MAR-2003.	
XX	PF	17-SEP-2002; 2002WO-US029560.	
XX	PR	17-SEP-2001; 2001US-0323469P.	
XX	PR	20-SEP-2001; 2001US-0323887P.	
XX	PR	13-NOV-2001; 2001US-0350666P.	
XX	PR	08-FEB-2002; 2002US-0355145P.	
XX	PR	08-FEB-2002; 2002US-0355257P.	
XX	PR	12-APR-2002; 2002US-0372246P.	
XX	PA	(BOSE-) EOS BIOTECHNOLOGY INC.	
XX	PI	Afar D, Aziz N, Gish KC, Hevezi PA, Mack DH, Wilson KB;	
XX	PI	Zlotnik A;	
XX	XX	WPI; 2003-354600/33.	
DR	DR	N-PSDB; ACC72821.	

CC	regulated in specific cancers (e.g. about 1031 genes up-regulated in
CC	acute lymphocytic leukemia). ACC72641 to ACC72860 represent cancer
CC	related gene nucleotide sequences which encode the proteins given in
CC	ABR58521 to ABR58709. Also described: (1) determining the presence or
CC	absence of a pathological cell in a patient; (2) an expression vector
CC	comprising a nucleic acid molecule described above; (3) a host cell
CC	comprising the vector; (4) an isolated polypeptide, which is encoded by
CC	the nucleic acid; (5) an antibody that specifically binds the polypeptide
CC	of (4); (6) specifically targeting a compound to a pathological cell in a
CC	patient by administering to the patient the antibody above; and (7) a
CC	drug screening assay. The nucleic acid is useful as diagnostic markers or
CC	therapeutic targets. In particular, the nucleic acid is useful for
CC	diagnosing a pathology, e.g. cancer (e.g. cancer of the bone marrow,
CC	bladder, brain, breast, cervix, colon/rectum, kidney, lung, ovary,
CC	pancreas, prostate, skin and uterus), wounds, ischaemia, heart diseases,
CC	atherosclerosis and endometriosis. The nucleic acid is also useful in
CC	drug screening, particularly for identifying agents for treating these
CC	pathologies
XX	
XX	
SQ	Sequence 829 AA;
	Query Match 99.9%; Score 4361.; DB 6; Length 829;
	Best Local Similarity 99.9%; Pred No. 0;
	Matches 829; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy	1 MGPRGPLASLLQLVCWLQCAASEPCRAVFREAVENTLEAGAEQEPGQALGVFMGCPG 60
Dd	1 MGPRGPLASLLQLVCWLQCAASEPCRAVFREAVENTLEAGAEQEPGQALGVFMGCPG 60
Qy	61 QEPALSTDNDDFTVNGETVQERRSIKERNPLKIPSKILRRHKRDWVAIVSPENG 120
Dd	61 QEPALSTDNDDFTVNGETVQERRSIKERNPLKIPSKILRRHKRDWVAIVSPENG 120
Qy	121 KGFPFQRLNQLKSNKDRTDKIFYSITPGADSPPEGVFAVEKETGMILLNKPLDREBIAK 180
Dd	121 KGFPFQRLNQLKSNKDRTDKIFYSITPGADSPPEGVFAVEKETGMILLNKPLDREBIAK 180
Qy	181 YELFGHAVSNGASVEDPNNISIIVTQNDHKKPKTQDTFRGSVLSEGLVPGTSVMQVAT 240
Dd	181 YELFGHAVSNGASVEDPNNISIIVTQNDHKKPKTQDTFRGSVLSEGLVPGTSVMQVAT 240
Qy	241 DEDDAIYTVNGVVAYSIHSQEPDPHDLMFTIHRSTGTISVISSGLDREKVPETYLTIA 300
Dd	241 DEDDAIYTVNGVVAYSIHSQEPDPHDLMFTIHRSTGTISVISSGLDREKVPETYLTIA 300
Qy	301 TMDMGSGSTTTAVAVVEILLDANDNAEMFDPOKYEAHVPENAVCHVEORLTVTDLDA NSP 360
Dd	301 TMDMGSGSTTTAVAVVEILLDANDNAEMFDPOKYEAHVPENAVCHVEORLTVTDLDA NSP 360
Qy	361 AWRAATYLIMGDDGDGHFTITTHPESNQGIITTRKGLDFEAKNQHTLYVEVTNEAPFVKL 420
Dd	361 AWRAATYLIMGDDGDGHFTITTHPESNQGIITTRKGLDFEAKNQHTLYVEVTNEAPFVKL 420
Qy	421 PTSTATIVHVDVNEAPVFPVPSKVVEVQEGIPTGEPCVCVTAEPDDKENQKISRILR 480
Dd	421 PTSTATIVHVDVNEAPVFPVPSKVVEVQEGIPTGEPCVCVTAEPDDKENQKISRILR 480
Qy	481 DPAGWLAMPDSQGVTAAGTLTDREDEQFVRNNIYEVWVWMLAMDNQSPPTTGCTFLLLTLID 540
Dd	481 DPAGWLAMPDSQGVTAAGTLTDREDEQFVRNNIYEVWVWMLAMDNQSPPTTGCTFLLLTLID 540
Qy	541 VNDHGVPBPFRQITICNQSPVRHLNITDKDSLPHTSPPQAQLTDDSDIYWTAENVBEGD 600
Dd	541 VNDHGVPBPFRQITICNQSPVRQLNITDKDSLPHTSPPQAQLTDDSDIYWTAENVBEGD 600
Qy	601 TVVLSLKXKLKQDYDVHLSLSDHGKNEQLTVIRATVCDCHGHVETCPGPWKGGFILPV L 660
Dd	601 TVVLSLKXKLKQDYDVHLSLSDHGKNEQLTVIRATVCDCHGHVETCPGPWKGGFILPV L 660
Qy	661 GAVIALLFLLLVLVLRKKRIKEPLLIPEDTRNVFYVREGGEGEDQDYDTIQLJHR 720
Dd	661 GAVIALLFLLLVLVLRKKRIKEPLLIPEDTRNVFYVREGGEGEDQDYDTIQLJHR 720

QY 721 GLEAPSEVLRNDVAPTIIPTEMYRPRPNDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 Db 721 GLEAPSEVLRNDVAPTIIPTEMYRPRPNDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 QY 781 DYEGSGDAASLSLTSSASDQDDYDLNWSRFFKFLADMYGGGDD 829
 Db 781 DYEGSGDAASLSLTSSASDQDDYDLNWSRFFKFLADMYGGGDD 829

RESULT 11
 ABUS6434
 ID ABUS6434 standard; protein; 829 AA.
 XX
 AC ABUS6434;
 XX
 DT 02-APR-2003 (first entry)
 XX
 DE Lung cancer-associated polypeptide #27.
 XX
 KW Lung cancer-associated polypeptide; cytostatic; emphysema;
 KW anti-inflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
 KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
 KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
 KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
 XX
 OS Unidentified.
 XX
 XX WO200286443-A2.
 PN
 XX
 PD 31-OCT-2002.
 XX
 XX 18-APR-2002; 2002WO-US012476.
 XX
 PR 18-APR-2001; 2001US-0284770P.
 PR 10-MAY-2001; 2001US-0250492P.
 PR 09-NOV-2001; 2001US-0339245P.
 PR 13-NOV-2001; 2001US-0350666P.
 PR 29-NOV-2001; 2001US-0334370P.
 PR 12-APR-2002; 2002US-0372246P.
 XX
 PA (EOSB-) EOS BIOTECHNOLOGY INC.
 XX
 PI Aziz N, Murray R;
 XX
 DR WPI; 2003-093161/08.
 DR N-PSDB; ABX76155.
 XX
 PT Detecting a lung cancer-associated transcript in a cell from a patient
 PT for treating lung cancer, by contacting a biological sample from the
 PT patient with a polynucleotide that exhibits increased or decreased
 PT expression in lung cancer.
 XX
 PS Claim 27; Page 210; 453pp; English.
 XX
 CC The invention relates to a method for detecting a lung cancer-associated
 CC transcript in a cell from a patient, comprising contacting a biological
 CC sample from the patient with a polynucleotide that selectively hybridizes
 CC to a sequence that is at least 80 % identical to a gene that exhibits
 CC increased or decreased expression in lung cancer samples. Lung cancer-
 CC associated polynucleotides and polypeptides are used for identifying a
 CC compound that modulates a lung cancer-associated polypeptide, for
 CC inhibiting proliferation of a lung cancer-associated cell to treat lung
 CC cancer in a patient and for treating a mammal having lung cancer by
 CC administering a modulatory compound identified. The methods are useful
 CC for treating lung cancer, such as small cell lung cancer, non-small cell
 CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,
 CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,
 CC hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and
 CC bronchiectasis. The genes, polynucleotides and polypeptides are useful
 CC for diagnostic purposes and as targets for screening for therapeutic
 CC compounds that modulate lung cancer, such as antibodies. Sequences
 CC ABUS6408-ABUS6745 represent lung cancer-associated polypeptides of the
 CC invention

XX Sequence 829 AA;
 SQ
 Query Match 99.8%; Score 4361; DB 6; Length 829;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MGLPRGPLASLLLLQVCWLQCAASEPCRAVFEAEVLEAGGAEQEPGQALGVFMGCPG 60
 Db 1 MGLPRGPLASLLLLQVCWLQCAASEPCRAVFEAEVLEAGGAEQEPGQALGVFMGCPG 60
 QY 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKDDWVAPISVPENG 120
 Db 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKDDWVAPISVPENG 120
 QY 121 KGPPQRLNQLKSNKDRDTKIFYISITGPADSPPEGVFAVEKETGWLKLNKPLDREIAK 180
 Db 121 KGPPQRLNQLKSNKDRDTKIFYISITGPADSPPEGVFAVEKETGWLKLNKPLDREIAK 180
 QY 181 YELFGHVSNGASVEDPMNISIIITQNDHKPKFTQDTFRGSVLGVLPCTSMQVAT 240
 Db 181 YELFGHVSNGASVEDPMNISIIITQNDHKPKFTQDTFRGSVLGVLPCTSMQVAT 240
 QY 241 DEDDAIYTYNGVAVSYHSQSPKDPHDLMTFTHSTGTISVISSGLDREKVPYTLTQIA 300
 Db 241 DEDDAIYTYNGVAVSYHSQSPKDPHDLMTFTHSTGTISVISSGLDREKVPYTLTQIA 300
 QY 301 TDMGCGSTTTAVAVVEILDANDNAPMFPDKYEAHVPENAVHVEVORLTVTDLDAENSP 360
 Db 301 TDMGCGSTTTAVAVVEILDANDNAPMFPDKYEAHVPENAVHVEVORLTVTDLDAENSP 360
 QY 361 AWRATYILMGDDGDGHFTITTHPSNQILTRKGLDFEAKNQHTLYVEVNEAPFVKL 420
 Db 361 AWRATYILMGDDGDGHFTITTHPSNQILTRKGLDFEAKNQHTLYVEVNEAPFVKL 420
 QY 421 PTSTATIVVHVVEDVNEAPVFPSPKVEVQEGITGPVPCVYTAEDPKENQKISYILR 480
 Db 421 PTSTATIVVHVVEDVNEAPVFPSPKVEVQEGITGPVPCVYTAEDPKENQKISYILR 480
 QY 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTLLTLLID 540
 Db 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTLLTLLID 540
 QY 541 VNDHGPVPEPQITTCNOSPVRHVLNITDKLSPTSPFOAQLTDDSDIYWTAEVNEEGD 600
 Db 541 VNDHGPVPEPQITTCNOSPVRHVLNITDKLSPTSPFOAQLTDDSDIYWTAEVNEEGD 600
 QY 601 TVVLSLKKFLKQDVTVDVHLSLSDHGKQELTVIRATVCDCHGHVETCPGPKGFIPLVL 660
 Db 601 TVVLSLKKFLKQDVTVDVHLSLSDHGKQELTVIRATVCDCHGHVETCPGPKGFIPLVL 660
 QY 661 GAVTALLFLLVLLVLRKKKIKPEPLLPEDDTRDNVYFYGEGGEGEDQDITQLHR 720
 Db 661 GAVTALLFLLVLLVLRKKKIKPEPLLPEDDTRDNVYFYGEGGEGEDQDITQLHR 720
 QY 721 GLEAPSEVLRNDVAPTIIPTEMYRPRPNDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 Db 721 GLEAPSEVLRNDVAPTIIPTEMYRPRPNDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 QY 781 DYEGSGDAASLSLTSSASDQDDYDLNWSRFFKFLADMYGGGDD 829
 Db 781 DYEGSGDAASLSLTSSASDQDDYDLNWSRFFKFLADMYGGGDD 829

RESULT 12
 ABUS6670
 ID ABUS6670 standard; protein; 829 AA.
 XX
 AC ABUS6670;
 XX
 DT 02-APR-2003 (first entry)
 XX
 DE Lung cancer-associated polypeptide #263.

KW	Lung cancer-associated polypeptide; cytostatic; emphysema;	QY	181	YELFGHASENGASVEDPMNIIIVTDQNDHHPKPTQDTFRGSVLEGLVPGTSMQVMTAT	240
KW	anti-inflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;	Db	181	YELFGHASENGASVEDPMNIIIVTDQNDHHPKPTQDTFRGSVLEGLVPGTSMQVMTAT	240
KW	small cell lung cancer; benign lesion; precancerous lesion; bronchitis;	QY	241	DEDDAIYVNGVWVAYSIIHQEPKPDHMLFTTHRSTGTISVSSGLDREKVPYETLTIOA	300
KW	chronic obstructive pulmonary disease; hypersensitivity pneumonitis;	Db	241	DEDDAIYVNGVWVAYSIIHQEPKPDHMLFTTHRSTGTISVSSGLDREKVPYETLTIOA	300
KW	interstitial pulmonary fibrosis; fibrosis; aschma; bronchiectasis.	QY	301	TDMDGGSTTTAVAVVEILLDANDNAPMPDPQYEAHVPENAVGHVQRLTVTDLDPNSP	360
XX	Unidentified.	Db	301	TDMDGGSTTTAVAVVEILLDANDNAPMPDPQYEAHVPENAVGHVQRLTVTDLDPNSP	360
XX	WO200286443-A2.	QY	361	AWRATYLMGDDGDHFTITTHPESNQGLITTRKGLDPEAKNQHTLYVEVNEAPFVLKL	420
XX	31-OCT-2002.	Db	361	AWRATYLMGDDGDHFTITTHPESNQGLITTRKGLDPEAKNQHTLYVEVNEAPFVLKL	420
XX	18-APR-2002; 2002WO-US012476.	QY	421	PTSTATIVVHVEDVNEAPVFPVPSKVVEQEGIPTEGVCVVTAEADPKENOKISYRILR	480
XX	18-APR-2001; 2001US-0284770P.	Db	421	PTSTATIVVHVEDVNEAPVFPVPSKVVEQEGIPTEGVCVVTAEADPKENOKISYRILR	480
XX	10-MAY-2001; 2001US-0290492P.	QY	481	DPAGLAMPDPSGQVTAAGTLDREDEQFVRNNIYEVWVLMNDNGSPPTTGTGLTLTLID	540
XX	09-NOV-2001; 2001US-0339245P.	Db	481	DPAGLAMPDPSGQVTAAGTLDREDEQFVRNNIYEVWVLMNDNGSPPTTGTGLTLTLID	540
XX	13-NOV-2001; 2001US-0350666P.	QY	541	VNDHGFVPEPROITICNQSPVRHVLNITDKOLSPHTSPFOQLTDDSDIYWTAEVNEEGD	600
XX	29-NOV-2001; 2001US-0334370P.	Db	541	VNDHGFVPEPROITICNQSPVRHVLNITDKOLSPHTSPFOQLTDDSDIYWTAEVNEEGD	600
XX	12-APR-2002; 2002US-0372246P.	QY	601	TVVLSLKFLKQDVTYVHLSLSDHGNKQOLTVIRATVCDCHGVETCPGPKGGFILPVL	660
XX	(EOSB-) EOS BIOTECHNOLOGY INC.	Db	601	TVVLSLKFLKQDVTYVHLSLSDHGNKQOLTVIRATVCDCHGVETCPGPKGGFILPVL	660
XX	Aziz N, Murray R;	QY	661	GAVALALLVLLVLLVLRKKRKKIPEPLLPEDDTDRDNVFFYEGEGGEBEDQDYDTQLHR	720
XX	WPI; 2003-093161/08.	Db	661	GAVALALLVLLVLLVLRKKRKKIPEPLLPEDDTDRDNVFFYEGEGGEBEDQDYDTQLHR	720
XX	N-PSDB; ABX76399.	QY	721	GLEARPEVVRNDVAPTIIPTWYPRPANDEIGNFIENLKAANTDPTAPPYDTLLVF	780
XX	Detecting a lung cancer-associated transcript in a cell from a patient	Db	721	GLEARPEVVRNDVAPTIIPTWYPRPANDEIGNFIENLKAANTDPTAPPYDTLLVF	780
XX	for treating lung cancer, by contacting a biological sample from the	QY	781	DYEGSGSDAASLSLTSSASDQDQDYDYLNEWGSFRFKLADMYGGGEDD	829
XX	patient with a polynucleotide that exhibits increased or decreased	Db	781	DYEGSGSDAASLSLTSSASDQDQDYDYLNEWGSFRFKLADMYGGGEDD	829
XX	expression in lung cancer.	XX	RESULT 13		
XX	Claim 27; Page 392; 453pp; English.	XX	AY09375		
XX	The invention relates to a method for detecting a lung cancer-associated	XX	ID	AY09375 standard; protein; 882 AA.	
XX	transcript in a cell from a patient, comprising contacting a biological	XX	XX	AY09375;	
XX	sample from the patient with a polynucleotide that selectively hybridises	XX	AC	AY09375;	
XX	to a sequence that is at least 80 % identical to a gene that exhibits	XX	DT	14-JUL-1999 (first entry)	
XX	increased or decreased expression in lung cancer samples. Lung cancer-	XX	DE	Wild-type E-cadherin.	
XX	associated polynucleotides and polypeptides are used for identifying a	XX	XX	E-cadherin; Maori; familial gastric cancer; germline mutation; detection;	
XX	compound that modulates a lung cancer-associated polypeptide, for	XX	KW	human; breast cancer; colorectal cancer; prostate cancer; thyroid cancer;	
XX	inhibiting proliferation of a lung cancer-associated cell to treat lung	XX	KW	kidney cancer; bladder cancer; liver cancer;	
XX	cancer in a patient and for treating a mammal having lung cancer by	XX	KW	hereditary diffuse gastric cancer; HDGC.	
XX	administering a modulatory compound identified. The methods are useful	XX	OS	Homo sapiens.	
XX	for treating lung cancer, such as small cell lung cancer, non-small cell	XX	XX	WO9920168-A2.	
XX	lung cancer or other benign or precancerous lesions, e.g. atelectasis,	XX	PN	29-APR-1999.	
XX	emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,	XX	XX	19-OCT-1998; 98WO-NZ000160.	
XX	hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and	XX	XX	17-OCT-1997; 97NZ-00328994.	
XX	bronchiectasis. The genes, polynucleotides and polypeptides are useful	XX	XX	(UYOT-) UNIV OTAGO.	
XX	for diagnostic purposes and as targets for screening for therapeutic	XX	PA	(TEWH-) TE WHETU WHANAU TRUST LTD.	
XX	compounds that modulate lung cancer, such as antibodies. Sequences				
XX	ABU56408-ABU56745 represent lung cancer-associated polypeptides of the				
XX	invention				
SQ	Sequence 829 AA;				
Query Match	99.6%; Score 4351; DB 6; Length 829;				
Best Local Similarity	99.8%; Pred.No. 0;				
Matches	827; Conservative 0; Mismatches 2; Indels 0; Gaps 0;				
QY	1 MGLPRGLASLLILQVCLQCAASEPCRAVFREAEVTLAAGAEQBPQALGVFMGCPG 60				
Db	1 MGLPRGLASLLILQVCLQCAASEPCRAVFREAEVTLAAGAEQBPQALGVFMGCPG 60				
QY	61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPFSKILRRHKRDWVAPISVPENG 120				
Db	61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPFSKILRRHKRDWVAPISVPENG 120				
QY	121 KGPPPPQRLNOLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREIAK 180				
Db	121 KGPPPPQRLNOLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREIAK 180				

QY 793 S S L T S S A S D Q Q D Y D Y L N E W G S R F K R L A D M Y G G E D D 829

Sequence 882 AA:

Query Match	55.4%	Score 2420	DB 4	Length 882
Best Local Similarity	55.1%	Pred. No. 2.2e-192		
Matches 483	Conservative 112	Mismatches 222	Indels	Gaps 13
QY	9	ASULLLQV-CWLQCAASBPCKRAVFEAEVLTLEAGGAQEPPGOALGV-FMGCPCG-QEPAAL	65	
DB	10	ALLULLQVSSWL-COEPEPCHPGDAESYFTVPRRHLEGRVILGRVNFEDCTGQRATY	68	
QY	66	PSTDNDDTTVNGETVQBRRLKERNP-----	92	
DB	69	FSLDT-REFKVGTDGVTYKRLRPHNPQIHFLVYAWDSTYRKFTKVTLTNTVGHHRPPP	127	
QY	93	-----LKTFF-SKRLRSHKDDWVAPISVPENKGPFPQORLNQKSNKDRDTK	140	
DB	128	HOASVSGTCAELLFFPNSPGLRQKQDWIPISCPENKGPFPKQVLQIKSNKDKGK	187	
QY	141	IFYSTGPGASPPGEGVFAVEKFGWLLLNKPLDRBEIAKYLEFHGAVSENGASVEDPMN	200	
DB	188	VFYSITGQCADTPVGVPIIBERTGWLKVTEPLDRERIATYTLFHAVSSNGNAVEDPME	247	
QY	201	ISGIIVTDQNDHKPKFTQDTPFGSLGVGLPGTSMQVTTATDEDDAITYLVNGVAYSHSQ	260	
DB	248	ILITVTDQNDKPKFTQEVFKGSWEGALPGTSMQVTTATDADDVNTYNAAIATYLSQ	307	
QY	261	EPKDPHDLMTFTHRSTGTISVISSGLDRKVPYTLTIQATMDGDSGTTTAAVAVEILD	320	
DB	308	DELPDKMFTINRNTGVISVVTTLGLDRSFPTTYLWQAADLQGEGLSTTATAVITVD	367	
QY	321	ANDNAPMEDPKQYEAHPENAVGHEVQRLVTDLDAPNSPAQRATVLLMGDGDGDHFTIT	380	
DB	368	TNDNPIIFNTPTYKGQVPEANEANVTTLKVTQDADAPNPAEAVTYIL-NDGQGFVVT	426	
QY	381	THPENOGILTRKGLDPEAKNQHTLVYVNEAPFLVTLPTSTATIVVHVEDVNEAPVF	440	
DB	427	TNPVNDGILTKAGLDPEAKQVILHVAVTNVVPEVSLTTSTATVTVDLVDVNEAPIF	486	
QY	441	VPPSKVVEVQGIPTGPEVCVYTAEDDK-ENOKISYRILRDPAGLWMDPDSQGVTAIG	499	
DB	487	VPEPKRVEVSDFGVGEISYTAQEPTFMQKITYRIWRDTANLWLEINPTGAISTRA	546	
QY	500	TLDRDEQFVANNIYVNMVLAAMDNGSPPTTGTTGLTLTLIDVNDHGPVPEPQITCNQS	559	
DB	547	ELDRDPEHVXNSIYALINATNGSPVATGTGLLLILSDVNDNAPIPEPRTIFFCERN	606	
QY	560	PVRHVLNTDKDLSPTSPFOAQLTDDSDIYWTAEVNE-EGTTVVLSLKFLQDQTYDVH	618	
DB	607	PKPQVINIADLPPNTSPFAELTHGASANWITQINDPTQESILTKPKMALEVGDYKIN	666	
QY	619	LSLSDHGNKBOLTIVRAVCDCHGVETC--PGFWKGF-----ILPVLGAVLALLFLVLV	672	
DB	667	LKLMNQNKQDVTILEVSVCDCEAGVCRKAQVENGLOI PAILGILGILALLILLL	726	
QY	673	LLLLVRKKRKIKEPILLPDEDDTRDNVPYEGEGGEDQDYITOLHRLGLEARPEVLURN	732	
DB	727	LLFLFRRRAVVKPEILLPPEDDTRDNVYYIDEGEGGEDQDFLSQHRGLDARPEVT-RN	785	
QY	733	DVAPTIITPTMYRPRPAPNDEIGNFIENLKAANTDPTAPPYDILLVFDYEGSGSDAASL	792	
DB	786	DVAPILMSVPRYLPRPAPNDEIGNFIENLKAANTDPTAPPYDILLVFDYEGSGSEASL	845	
QY	793	SSLTSSASDQDDYDYNLWNGSRFKKLADMYGGGEDD	829	
DB	846	SSLSNESDQDDYDYNLWNGENRFKKLADMYGGGEDD	882	

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:04:29 ; Search time 21 Seconds
(without alignments)

(without alignments)
2037.997 Million cell updates/sec

US-09-916-849A-1

Perfect score:

Sequence: 1 MGLPRGPLASLLQLVCWLQ.....NEWGSRFKKLADMYGGEDD 829

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

```

DO NOT REPRODUCE: Minimum Match 08

Post-processing:	Minimum	Match	0%
	Maximum	Match	100%

Maximum Match 100%
Testing first 45 summaries

Database : Issued Patents 22.*

```
Database : ISSUED Patents AA:
1. /can2 6/nt0Data/2/iaa/5A COMB ren.*
```

```

1: /cghz_6/pdata/2/1aa/3A_COMB.pdb:
2: /cghz_6/pdata/2/1aa/5B_COMB.pdb:

```

```
2: /cgn2_6/ptodata/2/1aa/3b_COMB.per:
3: /cgn2_6/ptodata/2/1aa/6a_COMB.per:*
```

```
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pdb: *
4: /cgn2_6/ptodata/2/1aa/6H_COMB.pdb: *
```

```
4: /cgn2_6/ptodata/2/1aa/6B_COMB.per:*
5: /cgn2_6/ptodata/2/1aa/6B_COMB.per:
```

```
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.del
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3553.5	81.3	822	2	US-08-474-067-9	Sequence 9, Appl
2	3553.5	81.3	822	2	US-08-474-068A-9	Sequence 9, Appl
3	3553.5	81.3	822	2	US-08-472-481-8	Sequence 8, Appl
4	2384	54.6	884	2	US-08-474-067-8	Sequence 8, Appl
5	2384	54.6	884	2	US-08-474-068A-8	Sequence 8, Appl
6	2384	54.6	884	2	US-08-472-481-7	Sequence 7, Appl
7	2348.5	53.8	878	1	US-08-237-912-2	Sequence 2, Appl
8	2348.5	53.8	878	4	US-08-733-429-2	Sequence 2, Appl
9	2348.5	53.8	878	4	US-09-798-267-3	Sequence 2, Appl
10	2348.5	53.8	878	4	US-09-798-267-3	Sequence 3, Appl
11	2348.5	53.8	878	5	PCR-US95-05518-2	Sequence 2, Appl
12	2231.5	51.1	837	2	US-08-474-067-7	Sequence 7, Appl
13	2231.5	51.1	837	2	US-08-474-068A-7	Sequence 7, Appl
14	2231.5	51.1	837	2	US-08-472-481-6	Sequence 6, Appl
15	1714	39.2	906	4	US-09-417-039-11	Sequence 11, Appl
16	1694	38.8	913	2	US-08-474-067-6	Sequence 6, Appl
17	1694	38.8	913	2	US-08-474-068A-6	Sequence 6, Appl
18	1694	38.8	913	2	US-08-472-481-5	Sequence 5, Appl
19	1621.5	37.1	916	1	US-08-188-228-48	Sequence 48, Appl
20	1621.5	37.1	916	1	US-08-332-643-42	Sequence 42, Appl
21	1621.5	37.1	916	1	US-08-332-638-48	Sequence 48, Appl
22	1175.5	26.9	555	2	US-08-453-702A-98	Sequence 98, Appl
23	1184	26.6	556	1	US-07-998-003A-98	Sequence 98, Appl
24	1184	26.6	556	1	US-08-453-724B-98	Sequence 98, Appl
25	1164	26.6	556	1	US-08-453-695A-98	Sequence 98, Appl
26	1164	26.6	556	1	US-08-268-161A-98	Sequence 98, Appl
27	1164	26.6	556	3	US-09-099-633-98	Sequence 98, Appl

Query Match 81.3%; Score 3553.5; DB 2; Length 822;
Best Local Similarity 81.5%; Pred. No. 0;
Matches 678; Conservative 58; Mismatches 83; Indels 13; Gaps 5;

1 MGI.PRGPLA-SLLLLVCWLOCAASEPCRAVF-REAETLEAGGAEQEPGQALGKVFMC 58

28	1023.5	23.4	713	1	US-08-188-228-62	Sequence 52, Appl
29	1023.5	23.4	713	1	US-08-332-643-56	Sequence 567, Appl
30	1023.5	23.4	713	1	US-08-332-633-62	Sequence 52, Appl
31	976	22.3	712	2	US-08-474-067-2	Sequence 2, Appl
32	976	22.3	712	2	US-08-474-067-5	Sequence 5, Appl
33	976	22.3	712	2	US-08-474-067-5	Sequence 2, Appl
34	976	22.3	712	2	US-08-474-068A-2	Sequence 5, Appl
35	976	22.3	712	2	US-08-474-068A-5	Sequence 2, Appl
36	976	22.3	712	2	US-08-472-481-2	Sequence 4, Appl
37	976	22.3	712	2	US-08-474-067-4	Sequence 4, Appl
38	976	22.3	717	2	US-08-474-068A-4	Sequence 4, Appl
39	968.5	22.2	796	2	US-08-738-349-4	Sequence 4, Appl
40	966.5	22.1	796	1	US-08-188-228-58	Sequence 58, Appl
41	966.5	22.1	796	1	US-08-332-643-52	Sequence 52, Appl
42	966.5	22.1	796	1	US-08-332-633-58	Sequence 58, Appl
43	951.5	21.8	796	2	US-08-738-349-2	Sequence 2, Appl
44	933	21.4	799	1	US-08-188-228-42	Sequence 42, Appl
45	933	21.4	799	1	US-08-332-638-42	Sequence 42, Appl

ALIGNMENTS

```

RESULT 1
US-08-474-067-9
; Sequence 9, Application US/08474067
; Patent No. 581158
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESS: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; City: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,067
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1682
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 822 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;
; US-08-474-067-9

```

Query Match	81.3%	Score 3553.5	DB 2	Length 822
-------------	-------	--------------	------	------------

```

Db 1 MELLSPHAFLLLLQVCLRSVSEVPYRAGFVLEVEGTDLEPSQVLGKVALAG 60
QY 59 PQQEPALFSTNDPFTVNGSTQVRSKERNPLKIFPSKILRRKRDWVAPISVPE 118
Db 61 QGMHHA-----DNGDIIMLTTRGTGQGDAMHS-----PPTRILRRKRWMPPIFVPE 110
QY 119 NGKGPFPORLQKSNKDRGTKIYFISITGPGADSPPEGVFAVEKETGWLKLLKPLDREBI 178
Db 111 NGKGPFPORLQKSNKDRGTKIYFISITGPGADSPPEGVFTIERESGWLLEMLDREKI 170
QY 179 AKYELFGHAVSENGASVEDPMNISIIITDQNDHFKPTQDTFRGSVLEGLPGTSVMQVT 238
Db 171 VKYELFGHAVSENGASVEEPMNISIIITDQNDHFKPTQDTFRGSVLEGLPGTSVMQVT 230
QY 239 ATDEDDALTYNGVAVYSHQEPKPDHMLFTTHRSTGTISVSSGLDREKVPYELTI 298
Db 231 ATDEDDAVNTYNGVAVYSHQEPKPDHMLFTTHKSTGTISVSSGLDREKVPYELTV 290
QY 299 QATMDGSGSTTTAVAVVEILDANDNAPMPDQKYEAHVPENAVGHEVQRLTVTDLDAPN 358
Db 291 QATMDGSGSTTTAAVAVVQILDANDNAPPEFQKYEAVWPNENEVQRLTVTDLDVFN 350
QY 359 SPANRATYLVINGDDGDHFTTHHPESNOGILLTRKGLDPEAKNOHTLYVEVTNEAPFVL 418
Db 351 WPANRATYLVINGDDGDHFTTHHPETNOGVLTTKGLDPEAQDHTLYVEVTNEAPFAV 410
QY 419 KLPTSTATIVVHVEDNEAPVFPVPSKVVEQEGIPTGEPVCVYTAEDPKENQKISYRI 478
Db 411 KLPTATATVAVVHVEDNEAPVFPVPSKVIEAQEGISIGELVCIVTAQDPKEDQKISYTI 470
QY 479 LRDPAGWLAMPDQSGQVTAAGTAVTGLDREDEQFVNNIYEVWVLMNDGSPPTTGTLLTL 538
Db 471 SRDPANWLAMPDQSGQVTAAGTAVTGLDREDEQFVNNIYEVWVLMNDGSPPTTGTLLTL 530
QY 539 IDVNDHGVPPEPRQITICNSQFVRHVLNITDKLSPTSFPFQALTDSDIYTAEVNEE 598
Db 531 TDINDHGPIPEPRQIICNSQFVPQVNLITDKLSPNSSPFQALTHSDIYWAASEK 590
QY 599 GDTVVLSSKKFLKQDVTVDVHLSLSDHGNKQRLTVIRATVCDCHGVH-ETCPGPKWGGFIL 657
Db 591 GDTVALSSKKFLKQDVTVDVHLSLSDHGNRQRLTVIRATVCDCHGVFNDQCPKWKGGFIL 650
QY 658 PVLGAVLALLLVLVLLVLRKKRKBEPILLPDDTRDNVYFYGGEGGEEBQDQYDIQ 717
Db 651 PILGAVLALLLVLVLLVLRKKRKBEPILLPDDTRDNVYFYGGEGGEEBQDQYDIQ 710
QY 718 LHRGLEARPEVLRNDVAPTIIPPMYRPRANPDEIGNFIENLKAANTDPTAPPYDIL 777
Db 711 LHRGLEARPEVLRNDVAPTIIPPMYRPRANPDEIGNFIENLKAANTDPTAPPYDSL 770
QY 778 LVPYEGSGDAASLSISLTSASQDQDXYDLNWSGRFEXKLADMYGGGDD 829
Db 771 MVFDYEGSGDAASLSLTSASQDQDXYDLNWSGRFEXKLADMYGGGDD 822

```

RESULT 2

```

US-08-474-068A-9
; Sequence 9, Application US/08474068A
; Patent No. 5837525
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,068A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994
; PRIOR APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1683
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 822 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-474-068A-9

```

```

Query Match 81.3%; Score 3553.5; DB 2; Length 822;
Best Local Similarity 81.5%; Pred. No. 0;
Matches 678; Conservative 58; Mismatches 83; Indels 13; Gaps 5;

```

```

QY 1 MGLPGLA-SILLQVCLQCAASEPCRAVP-REAEVTLAAGAEQEPGQALGKVMGC 58
Db 1 MELLSPHAFLLLLQVCLRSVSEVPYRAGFVLEVEGTDLEPSQVLGKVALAG 60
QY 59 PQQEPALFSTNDPFTVNGSTQVRSKERNPLKIFPSKILRRKRDWVAPISVPE 118
Db 61 QGMHHA-----DNGDIIMLTTRGTGQGDAMHS-----PPTRILRRKRWMPPIFVPE 110
QY 119 NGKGPFPORLQKSNKDRGTKIYFISITGPGADSPPEGVFAVEKETGWLKLLKPLDREBI 178
Db 111 NGKGPFPORLQKSNKDRGTKIYFISITGPGADSPPEGVFTIERESGWLLEMLDREKI 170
QY 179 AKYELFGHAVSENGASVEDPMNISIIITDQNDHFKPTQDTFRGSVLEGLPGTSVMQVT 238
Db 171 VKYELFGHAVSENGASVEEPMNISIIITDQNDHFKPTQDTFRGSVLEGLPGTSVMQVT 230
QY 239 ATDEDDALTYNGVAVYSHQEPKPDHMLFTTHRSTGTISVSSGLDREKVPYELTI 298
Db 231 ATDEDDAVNTYNGVAVYSHQEPKPDHMLFTTHKSTGTISVSSGLDREKVPYELTV 290
QY 299 QATMDGSGSTTTAVAVVEILDANDNAPMPDQKYEAHVPENAVGHEVQRLTVTDLDAPN 358
Db 291 QATMDGSGSTTTAAVAVVQILDANDNAPPEFQKYEAVWPNENEVQRLTVTDLDVFN 350
QY 359 SPANRATYLVINGDDGDHFTTHHPESNOGILLTRKGLDPEAKNOHTLYVEVTNEAPFVL 418
Db 351 WPANRATYLVINGDDGDHFTTHHPETNOGVLTTKGLDPEAQDHTLYVEVTNEAPFAV 410
QY 419 KLPTSTATIVVHVEDNEAPVFPVPSKVVEQEGIPTGEPVCVYTAEDPKENQKISYRI 478
Db 411 KLPTATATVAVVHVEDNEAPVFPVPSKVIEAQEGISIGELVCIVTAQDPKEDQKISYTI 470
QY 479 LRDPAGWLAMPDQSGQVTAAGTAVTGLDREDEQFVNNIYEVWVLMNDGSPPTTGTLLTL 538
Db 471 SRDPANWLAMPDQSGQVTAAGTAVTGLDREDEQFVNNIYEVWVLMNDGSPPTTGTLLTL 530
QY 539 IDVNDHGVPPEPRQITICNSQFVRHVLNITDKLSPTSFPFQALTDSDIYTAEVNEE 598
Db 531 TDINDHGPIPEPRQIICNSQFVPQVNLITDKLSPNSSPFQALTHSDIYWAASEK 590
QY 599 GDTVVLSSKKFLKQDVTVDVHLSLSDHGNKQRLTVIRATVCDCHGVH-ETCPGPKWGGFIL 657
Db 591 GDTVALSSKKFLKQDVTVDVHLSLSDHGNRQRLTVIRATVCDCHGVFNDQCPKWKGGFIL 650

```

QY 658 PVLGAVLALLFLLVLLLVKRRKIKKEPILLPEDDTRDNVFXGEGGSEDDQDYDITQ 717
Db 651 PIUGAVLALLTLLALLLVKRRKIKKEPILLPEDDTRDNVFXGEGGSEDDQDYDITQ 710
QY 718 LHRGLEARPEVLRNDVAPTIPTMYRPRPANPDEIGNFIENLKAANTDTPAPPYDTL 777
Db 711 LHRGLEARPEVLRNDVAPTIPTMYRPRPANPDEIGNFIENLKAANTDTPAPPYDSL 770
QY 778 LVFDYEGSGSDAASLSLTSASDQDYDYLNEWGSRFKKLADMYGGGEDD 829
Db 771 MVFDYEGSGSDAASLSLTSASDQDYDYLNEWGSRFKKLADMYGGGEDD 822

RESULT 3
US-08-472-481-8
; Sequence 8, Application US/08472481
; Patent No. 5863804
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,481
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994
; PRIOR APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1686
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 822 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-472-481-8

Query Match 81.3%; Score 3553.5; DB 2; Length 822;
Best Local Similarity 81.5%; Pred. No. 0;
Matches 678; Conservative 58; Mismatches 83; Indels 13; Gaps 5;

QY 1 MGLPRGFLA-SLLLVQVLCQAAEPCRAVF-REARVTLAAGAEQPGQALGKVFMC 58
Db 1 MELLGPHAFILLLVQVLCWLSVSEPYRAGFVGEAGVTLVETGDLFSPQVKGVALAG 60
QY 59 PGQPALFSTDNDFTVRNGETVQERSLKERNPDKIFPSKRLRHRKRDVWVAPISVPE 118
Db 61 QGMEHA---DNGDIIMLTGRVQGGKDAHMS-----PPTRILRRKREWVMPPIFVPE 110
QY 119 NGKGPFFORLNOLKSNDRKTKIFYSITGCA DSPPEGVFAVEKCTGWLINLPDSEET 178
Db 111 NGKGPFFORLNOLKSNDRGRTKIFYSITGCA DSPPEGVFTIEKESGWLINLPDREKI 170

QY 179 AKYELFGHAVSENGASVEDPMNIIIVTDQNDHKPKFTQDTRFGSVLEGLVPGTSVMQVT 238
Db 171 VKYELGHAVSENGASVEEPMNIIIVTDQNDNPKFTQDTRFGSVLEGVMPGTSVMQVT 230
QY 239 ATDEDDALITYNGVAVSIHSQEPKPDHDMFTTHRSTGTSTISVSSGLDRKKEVPEYILT 298
Db 231 ATDEDDAVNTYNGVAVSIHSQEPKPDHDMFTTHKSTGTSTISVSSGLDRKKEVPEYILT 290
QY 299 QATMDGSGSTTAVVVEILLDANDNAPMFPDQKYEAHVHPENAVGHEVQRILTVDLDPN 358
Db 291 QATMDGSGSTTAEAVVQILLDANDNAPFEPFQKYEAVHPENAVGHEVQRILTVDLDPN 350
QY 359 SPAMRATYLMGGDDGHFTITTHPEQNQGLITRKLGLDFEAKNQHTLYVEVTNEAPFVL 418
Db 351 WPAWRATYHVGDDGDGHFTITTHPETNQGLITRKLGLDFEAKNQHTLYVEVTNEAPFV 410
QY 419 KLPTSTATIVVHVDVNEAPVFPVPSKVVEQGIPTGEPVCVYTAEDPDKENOKISYRI 478
Db 411 KLPTATATVVVHVVDVNEAPVFPVPSKVVEQGISIGELVCIYTAQDPDKEDQKISYTI 470
QY 479 LRDPAWMLAMPDSGQVTAAGTLDREDEQFVKNVYVWVWLATDSGNPPTTGTGTLTL 538
Db 471 SRDPAWMLAVDPDSQITAAAGILDREDEQFVKNVYVWVWLATDSGNPPTTGTGTLTL 530
QY 539 IDVNDHGPVPEPRITICNQSPVRHVLNITDKDLSPTSPFOAQLTDDSDIYWTAEVNEE 598
Db 531 TDINDHGPPEPRQIIICNQSPVQVLNITDKDLSPNSSPFOAQLTDDSDIYWTAEVSEK 590
QY 599 GDTVVLKXFLKQDVTYVHLSLSDHGNKEQLTWIRATVCDCHGVH-ETCPGPKWGGFIL 657
Db 591 GDTVALSLKXFLKQDVTYDLHLSLSDHGNREQLTWIRATVCDCHGVFNDGPRPKWGGFIL 650
QY 658 PVLGAVLALLFLLVLLLVKRRKIKKEPILLPEDDTRDNVFXGEGGSEDDQDYDITQ 717
Db 651 PIUGAVLALLTLLALLLVKRRKIKKEPILLPEDDTRDNVFXGEGGSEDDQDYDITQ 710
QY 718 LHRGLEARPEVLRNDVAPTIPTMYRPRPANPDEIGNFIENLKAANTDTPAPPYDTL 777
Db 711 LHRGLEARPEVLRNDVAPTIPTMYRPRPANPDEIGNFIENLKAANTDTPAPPYDSL 770
QY 778 LVFDYEGSGSDAASLSLTSASDQDYDYLNEWGSRFKKLADMYGGGEDD 829
Db 771 MVFDYEGSGSDAASLSLTSASDQDYDYLNEWGSRFKKLADMYGGGEDD 822

RESULT 4
US-08-474-067-8
; Sequence 8, Application US/08474067
; Patent No. 5811518
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,067
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1682
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-474-067-8

Query Match 54.6%; Score 2384; DB 2; Length 884;
Best Local Similarity 53.5%; Pred. No. 7e-213;
Matches 471; Conservative 123; Mismatches 221; Indels 66; Gaps 13;

QY 9 ASLLLLQV-CWLQCAASEP--CRAVFREAETLEAGAGOEFGOALGKV-FMGCPCQGEPA 64
DB 10 ALLLLQVSSWL-COELEPSCSPGFSSEVYFPVPERHLERGHVLRGVFEGCTGRPT 68
QY 65 LFSTDNDFTVRNGETVQERRSLK----- 88
DB 69 AFFSEDSRFKATDGTITVKEHLKHLKLETSFLVRAARDSSHRELSTKVTLKSMGHHHRH 128
QY 89 -----ERNP-LKIFPSKRI-LRRHKRDVWVAPIVSPENKGGFFPQRLNQLKSNKDRDT 139
DB 129 HHRDPAGESNELLMFFSVPLGRQRKRDWIPPIPCPENEKGEFPKNLVQIKSNRDKET 188
QY 140 KIFYSIITGPAGDSPPGEGVFAVEKETGMLLNKPLDREIAKYELFGHVSNGEASVEDPM 199
DB 189 KVFYSITGQAGDKPVGVIETRTGWLKVTQPLDRAIAKYILYSHAVSSNGEAVEDPM 248
QY 200 NISIVTQDNHDKPKFTQDTRFSGVLEGLVPGTSMQVATDDEDAIYTYNGVAVSIHS 259
DB 249 EIVITVTDQDNREFTQEVFEGSVAREGAVPGTSMKVSATDADDDVNTYNAIAYTVS 308
QY 260 QEPKDPHLMFTIHRSGTISVSSGLDREKVPETLTIOATMDGDSGTTTAVAVVEIL 319
DB 309 QDPPLPHKNMFTVNRDTGVSVLTSGLDRESYTYTLVQADLQSEGSLTKAKAVITVK 368
QY 320 DANDNAPMFPQKYEAHVPENAVGEVQRLTVTDLDPNPSAPWATYILMGSGDGDHFTI 379
DB 369 DINDNAPVFNFPSTYQGGVPENEVNARIATLKVTDGDDAPNTPAMKVVTYV-NDPDCQFVV 427
QY 380 TTPESNOGILTRKGLDFAKQOHTLYVEVTNEAPVCLKPTSTATIVVHVEDVNEAPV 439
DB 428 VTDDTTNDGILKTAGGLDFAKQYILHVRVNEEPEPESGLVSTATVTVVVDVNEAPI 487
QY 440 FVPEKVVYQEGIPTGPEVCVYTAEDPK-ENQKISYRILRDPAGWLAMPDPSGQVAV 498
DB 488 FMPAERSEVPEDFGVQGEITSYTAREPDTFMDQKITYRWRDTANWLEINPETGAI 547
QY 499 GTLDREDEQFRNNIYVWVLMDNGSPPTTGCTGLLLTLIDVNDHGPVPEPQITICNQ 558
DB 548 AEMDREDAEHVKNSTYVALIATDGGSPATGTTLLVLLDNDNAPPEPENNMOFCOR 607
QY 559 SPVRHVLNITDKOLSPHTSPFQALTDSDIYVTAEVNBEQ-DTVVLSLKKFLKQDTYDV 617
DB 608 NPQPHIITILDPLDPPNTSFTAEHTHGSVNTIENDAQAQSSLLIQPKOLEIGEYKI 667
QY 618 HLSLSDHGNKEQLTVIRATVCDCHGVHETCPGPKWGGF-----ILPVIGAVLALLF 668
DB 668 HLKADQNQKQVTVTLVDVHVDCDEGTVNNC--MKAGIVAAGLQVPAIILGILGILALI 724
QY 669 LIALVLLVLRKKIKIPEPLLPDDTDRNVYVYGBEGGEDDQDYDITOLHGLEARPEV 728
DB 725 LIALVLLVLRKKIKIPEPLLPDDTDRNVYVYGBEGGEDDQDYDITOLHGLEARPEV 784

```

```

QY 729 VLRNDVAPTIITPTMYRPRPANPDEIGNFIENLKAANTDPTAPPYDTLLVFDYEGSGSD 788
DB 785 T-RNDVAPTLMSVPOYRPRPANPDEIGNFIENLKAADSDPTAPPYDLSLLVFDYEGSGSE 843
QY 789 AASLSLSSASDQDDYDYLNEWGSRFKKLADMYGGGEDD 829
DB 844 AASLSLSSASDQDDYDYLNEWGSRFKKLADMYGGGEDD 884

RESULT 5
US-08-474-068A-8
; Sequence 8, Application US/08474068A
; Patent No. 5837525
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/474, 068A
; APPLICATION NUMBER: US/08/474, 068A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1683
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-474-068A-8

```

```

Query Match 54.6%; Score 2384; DB 2; Length 884;
Best Local Similarity 53.5%; Pred. No. 7e-213;
Matches 471; Conservative 123; Mismatches 221; Indels 66; Gaps 13;

QY 9 ASLLLLQV-CWLQCAASEP--CRAVFREAETLEAGAGOEFGOALGKV-FMGCPCQGEPA 64
DB 10 ALLLLQVSSWL-COELEPSCSPGFSSEVYFPVPERHLERGHVLRGVFEGCTGRPT 68
QY 65 LFSTDNDFTVRNGETVQERRSLK----- 88
DB 69 AFFSEDSRFKATDGTITVKEHLKHLKLETSFLVRAARDSSHRELSTKVTLKSMGHHHRH 128
QY 89 -----ERNP-LKIFPSKRI-LRRHKRDVWVAPIVSPENKGGFFPQRLNQLKSNKDRDT 139
DB 129 HHRDPAGESNELLMFFSVPLGRQRKRDWIPPIPCPENEKGEFPKNLVQIKSNRDKET 188
QY 140 KIFYSIITGPAGDSPPGEGVFAVEKETGMLLNKPLDREIAKYELFGHVSNGEASVEDPM 199
DB 189 KVFYSITGQAGDKPVGVIETRTGWLKVTQPLDRAIAKYILYSHAVSSNGEAVEDPM 248

```

```
QY 200 NISIIIVTDNDHKPKFTQDFRGSVLEGLVPGTSVMQVATDDEDDALITYNGVVAYSIHS 259
Db 249 EIVITVTDQNDNRPEFTQEFVGSVAEGVGSWMKVSATDADDVNTVNAIATIVS 308
QY 260 QEPKOPHLMFTIHRSTGTTSVSSGLDREKVPYTLTIQATDMGDSGTTTAVAVVEIL 319
Db 309 QDPPLPHKNMFTVNRDTGVSVLTSLDRESYPTYLIVVQAADLQEGSLSTAKAVITVK 368
QY 320 DANDNAPMFDPOKYEAHVPENAVCHEVQRLVTVDLDPNSPAMRATYILMGDDGDBHFTI 379
Db 369 DINDNAPVFNSTYQGVPEVNEVARIATLVKVTDDAPNTPAWKVVTYV-NDPDQGFV 427
QY 380 TTHPESNOGILTRKGLDFAKQHTLYVEVNEAPFLKLTSTATIVVHVEDVNEAPV 439
Db 428 VTDPTNDGILKTAAGLDFAKQYILHVRVENEPEPESGLVPSTATVTVVVDVNEAPI 487
QY 440 FVPSKVVEVQEGIPTGEPVCVTAEDDPK-ENQKISYRILRDPAGWLAMPDPSGQVAV 498
Db 488 FMPAERRVEVPEDFGVQGEITSYTAREPDTFMDQKITVIRWRDTANWLEINPETAIFTR 547
QY 499 GTLDREDEQFVRNNIYEVWVLAMONGSPPTGTGTLTLLTLDVNDHGPVPEPQITICNQ 558
Db 548 AEMDREDAEHVKNSTYVALIATDDGSPATGTGTLTLLVLLVLDVNDNAPPEPNWQFCOR 607
QY 559 SPVRHLNITDKLSPTSPTAQLTDDSDIYTAEVNEEG-DTVVLSLKKFLKQDTYDV 617
Db 608 NPQPHIITILDPLPNTSPFTAEITHGASVNWITIEYNDAAQESLILQPRKOLEIGEYKI 667
QY 618 HLSLSDHGNKQOLTVIRATVCDCHGHVETCPGPKWGGF-----ILPVLGAVLALLF 668
Db 668 HLKLDNQNDQVTTLDVHVDCGEGTVNNC---MKAGIVAAGLQVPAIILGILGILALLI 724
QY 669 LLLVLLLVKRRKIKIPELLLPEDDTRDNVYFYGEGGEGEDQDYDITQLHRLGEARPEV 728
Db 725 LILLLLFLRRRTVVKPELLPPDDTRDNVYFYGEGGEGEDQDFLSQLHRLGDARPEV 784
QY 729 VLNDVAPTIIPMPYRPRANPDEIGNFIENLKAANTDPTAPPYDILLVFDYEGSGSD 788
Db 785 T-RNDVAPTLMSVQVPRPRANPDEIGNFIENLKAANTDPTAPPYDILLVFDYEGSGSD 843
QY 789 AASLSSLTSSASQDQDYDYLNEWGSFRKFLADMYGGEGEDD 829
Db 844 AASLSSLNSESQDQDYDYLNEWGNFRKFLADMYGGEGEDD 884
```

RESULT 6

```
US-08-472-481-7
; Sequence 7, Application US/08472481
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,481
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1686
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 884 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-472-481-7
```

Query Match 54.6%; Score 2384; DB 2; Length 884;

Best Local Similarity 53.5%; Pred. No. 7e-213;

Matches 471; Conservative 123; Mismatches 221; Indels 66; Gaps 13;

```
QY 9 ASLLLLQV-CWLQCAASEP--CRAVFREAETLEAGGAEQPGQALGV-FWGCGFGQBP 64
Db 10 ALLLLLLQVSSWL-CQELPESCSFGSSEVYTFVPVPERHLERGHVLRGVRFGCTGRPT 68
QY 65 LFTSNDNFTVRNGSTVQERRSLK----- 88
Db 69 AFPSDSEFKVATDGTITVKGRLHLKLETSFLVZARDSSHRELSTKVTLSKMGHHRH 128
QY 89 -----SRNP-LKTFPSKRI-LRRHKEDWVAPISVPEKNGKPPORLNOLKSNKRD 139
Db 129 HRDPASNSPELLMFPSPYGLRRQKRDWIPPLISCENKGGFPKMLVQIKSNRDET 188
QY 140 KIFYSITGFGADSPPEGVFAVEKETGWLLLNKPLDREIAKYELFCHAVSENGASVEDPM 199
Db 189 KVFYSITGQAGKPPGVGVFIETRETGWLKVTPLDREAIKAYILYSHAVSSNGEAVEDPM 248
QY 200 NISIIIVTDNDHKPKFTQDFRGSVLEGLVPGTSVMQVATDDEDDALITYNGVVAYSIHS 259
Db 249 EIVITVTDQNDNRPEFTQEFVGSVAEGVGSWMKVSATDADDVNTVNAIATIVS 308
QY 260 QEPKOPHLMFTIHRSTGTTSVSSGLDREKVPYTLTIQATDMGDSGTTTAVAVVEIL 319
Db 309 QDPPLPHKNMFTVNRDTGVSVLTSLDRESYPTYLIVVQAADLQEGSLSTAKAVITVK 368
QY 320 DANDNAPMFDPOKYEAHVPENAVCHEVQRLVTVDLDPNSPAMRATYILMGDDGDBHFTI 379
Db 369 DINDNAPVFNSTYQGVPEVNEVARIATLVKVTDDAPNTPAWKVVTYV-NDPDQGFV 427
QY 380 TTHPESNOGILTRKGLDFAKQHTLYVEVNEAPFLKLTSTATIVVHVEDVNEAPV 439
Db 428 VTDPTNDGILKTAAGLDFAKQYILHVRVENEPEPESGLVPSTATVTVVVDVNEAPI 487
QY 440 FVPSKVVEVQEGIPTGEPVCVTAEDDPK-ENQKISYRILRDPAGWLAMPDPSGQVAV 498
Db 488 FMPAERRVEVPEDFGVQGEITSYTAREPDTFMDQKITVIRWRDTANWLEINPETAIFTR 547
QY 499 GTLDREDEQFVRNNIYEVWVLAMONGSPPTGTGTLTLLTLDVNDHGPVPEPQITICNQ 558
Db 548 AEMDREDAEHVKNSTYVALIATDDGSPATGTGTLTLLVLLVLDVNDNAPPEPNWQFCOR 607
QY 559 SPVRHLNITDKLSPTSPTAQLTDDSDIYTAEVNEEG-DTVVLSLKKFLKQDTYDV 617
Db 608 NPQPHIITILDPLPNTSPFTAEITHGASVNWITIEYNDAAQESLILQPRKOLEIGEYKI 667
QY 618 HLSLSDHGNKQOLTVIRATVCDCHGHVETCPGPKWGGF-----ILPVLGAVLALLF 668
Db 668 HLKLDNQNDQVTTLDVHVDCGEGTVNNC---MKAGIVAAGLQVPAIILGILGILALLI 724
QY 669 LLLVLLLVKRRKIKIPELLLPEDDTRDNVYFYGEGGEGEDQDYDITQLHRLGEARPEV 728
Db 725 LILLLLFLRRRTVVKPELLPPDDTRDNVYFYGEGGEGEDQDFLSQLHRLGDARPEV 784
QY 729 VLNDVAPTIIPMPYRPRANPDEIGNFIENLKAANTDPTAPPYDILLVFDYEGSGSD 788
```

```

Db      785 T-RNDVAPLMSVQXRPANDEIGNFIDENLKAADSDPTAPPYDLSLLVDFEGSGSE 843
Qy      789 AASLSLTSASDQDQDYDLNWSGSRFKKLADMYGGGDD 829
Db      844 AASLSLSSSDQDQDYDLNWSGSRFKKLADMYGGGDD 884

RESULT 7
US-08-237-919-2
; Sequence 2, Application US/08237919
; Patent No. 5610281
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael B
; APPLICANT: Cepak, Karyn L
; TITLE OF INVENTION: Methods and Compositions for
; Modulating Heterotypic E-cadherin Interactions with T lymphocytes
; TITLE OF INVENTION: Modulating Heterotypic E-cadherin Interactions with T lymphocytes
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESS: Wolf, Greenfield & Sacks P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/237,919
; FILING DATE:
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 878 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-237-919-2

Query Match      53.8%; Score 2348.5; DB 1; Length 878;
Best Local Similarity 54.0%; Pred. No. 1.4e-209;
Matches 472; Conservative 115; Mismatches 230; Indels 57; Gaps 12;

Qy      8 LASLLLOVQLQCAAEPCRAVREAEVLEAGAEQBPQALGKV-FMGCPGQZ----- 62
Db      8 LSGLLLLRSPLGQERSPPCLTRELHVGAGAPPEKRP--RLGRVNFEDCTGQRQTAI 65
Qy      63 -----PALFSTDNDQDFTVRNGETVQERRSLK-----ERNP--- 92
Db      66 FLTPPKVGTGDTVTRKPRNFENPTDPLGLRWDSYRFSKVTINTVGHHRPPPHQ 125
Qy      93 -----LKIFF-SKRLRRHRDVRVAPISVPENKGPFPFQRLNQLKSNKDRDTKIF 142
Db      126 ASVSGIQAEALLTFNSPGLRRQRKQDWIPPIPCPENKGPFPFQRLNQLKSNKDRDTKIF 185
Qy      143 YSITCPGADSPPEGVFAVEKTEGHLNKLKPLDREIAKVELFGHAYSENGASVEDPMNIS 202
Db      186 YSITGQGGADTPVGVFIETRTGKTEPLDRERATYTLFHAVSSNGNAVEDPMNELL 245
Qy      203 IIVTDQNDHKPKFTQDTFRGSLVGLPGTSVMQVATDDEDAIYTVNGVAVYSIHSQEP 262
Db      246 ITVTDQNDNKEFTEQVEFKGSMGALPGTSVMEVATDADDDVNTYNAIAVTLISQDP 305

```

```

Qy      263 KDPHDLMTIHRSTGTSIVSSGLDREKVPYTLTIQATMDMGSGSTTTAVAVVEILLDAN 322
Db      306 ELFDKDMFTINRNTGIVSVVTTGLDRESFPTYTLVQAAADIQSGSLSTTAVAVITVDTN 365
Qy      323 DNAPMFDQKYEAHVPENAVCHVEQRLTVTDLDAPNSPAMRATYLINGGDDGDFHTITTH 382
Db      366 DNPPIFNFTYKGVPEANEANVITTLKVTDDADAPNPAWEAVYIIL-NDGGGFVVITIN 424
Qy      383 PESNQILITRKGLDFFAKNQHTLYVEVNEAPFVCLKLPTSTATIVVHVVEDVNEAPFVFP 442
Db      425 PVNNDGILKTAAGLDFFAKQYILHVAVTNVVPEVSLTTSTATVTVDVLVDVNEGPVFP 484
Qy      443 PSKVVEOEGIPTGEPVCVYTAEDPK-ENOKISYRILRDPAGWLAMPDPSQGVTAVTGL 501
Db      485 PEKRVESSEDFGVGQETSYTAQSPDTFMEQKITYRIRWRDTRNWLEINPDGTGAISTRAEL 544
Qy      502 DREBQFVRNNIYEMVWLAMNGSPPTGTGTLLTLIDVNDHGVPEPEPQITTCNQSPV 561
Db      545 DREDFEVRNNTYALIIATDNGSPVATGTGTLILLSDVNDNAPIPEPRITPFCERNPK 604
Qy      562 RHLNITDKLSMTPSPQAOLTDSDIYTAEVNE-EGDTVLSLKKFLKQDTYDVHLS 620
Db      605 PQVINIHDAFLPNTSPFTELTGHRVFNWITIQNDPTQESIIILKPKMALEVGDFKINLK 664
Qy      621 LSDHGNKEQLTIRATVCDCHGVETC--PGPMKGGF---ILPVLGAVLALLFLLLVLL 674
Db      665 LMDNQKQVTTLEVSVCDCGAGVCRKAQPVFAGLQIPAILGILGILALLLILL 724
Qy      675 LLVRKKRIKEPLLPEDDTRDNVYFEGEGGEDQDYDTQLHRLGLEARPEVVLNDV 734
Db      725 LFLRRRAVVKPELLPPEDDTRDNVYFEGEGGEDQDYDTQLHRLGLEARPEVVLNDV 783
Qy      735 APTIPTMYRPRANDEIGNFIENKAANTDPTAPPYDLSLLVDFEGSGSDAASLSS 794
Db      784 APTLMSVPRYLRPANPDEIGNFIDENLKAADTPTAPPYDLSLLVDFEGSGSDAASLSS 843
Qy      795 LTSSASDQDQDYDLNWSGSRFKKLADMYGGGDD 828
Db      844 LNSSESXQDQDYDLNWSGSRFKKLADMYGGGDD 877

RESULT 8
US-08-732-429-2
; Sequence 2, Application US/08732429
; Patent No. 6300080
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic
; E-cadherin Interactions with T Lymphocytes
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Wolf, Greenfield & Sacks P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/732,429
; FILING DATE: herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/237,919
; FILING DATE: 3 May 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7023
; TELECOMMUNICATION INFORMATION:

```

TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 878 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-732-429-2

Query Match 53.8%; Score 2348.5; DB 4; Length 878;
Best Local Similarity 54.0%; Pred. No. 1.4e-209;
Matches 472; Conservative 115; Mismatches 230; Indels 57; Gaps 12;

YQ 8 LASLLQVLCWCAASEPCRAVFEAEVTLFAGGAEQEPQALGKV-FMGCPCGQ-----62
DB 8 LSGLLLLRSPLGQERSPPCLTRFLHVGAPAPPEKRP--RLGRVNFEDCTGQRTAI 65
YQ 63 -----PALFSTDNDDFTVRNGETVQERRSLK-----ERNP-----92
DB 66 FLTPIPKVGTDGVTIKRPLRFHNPDPFLGLRWDSYRKFSKVTLTNTVGHHRPPPHQ 125
YQ 93 -----LKIIP-SKRIIRHRKRDWVAPISVPENGKGFPPQRLNOLKSKNDRTKIF 142
DB 126 ASVSGIQAELLTFPNSPGLRRQRDWIPIPCPENEGKGFPPKNLVQKSKNKEGKVF 185
YQ 143 YSITGPGADSPPEGVFAVEKETGMLLNKPLDREEIAKYELFCHAVSENGASVEDPMNIS 202
DB 186 YSITGQADTPPVGVFIERTGMLKVTPELDREARIATYTLFHAVSSNGNAVEDPWEIL 245
YQ 203 IIVTDQNDHKPKTQDTPFGSVLEGVLPGTSVMQVATDEDDAIYTYNGVVAISHQEP 262
DB 246 IIVTDQNDKPEFTQEVFKGSVNEGALPGTSVMEVATDADDVNTYNAAIATYILSDP 305
YQ 263 KDPHLMFTIHRSTGTSVISGLDREKVPETLTQATDMGDGSGTTTAVAVVEILDAN 322
DB 306 ELDPKWMFTINRNTGVISVVTGLDRESFPTTLVQVQADLQGEGLSTATAVITVTDN 365
YQ 323 DNAPMFDPOKYEAHVPENAVCHVEQRLTVTDLPNPSPAWRATYILMGDDGDGHTTTH 382
DB 366 DNPPIFNPTTKYQVPEANVITLKVTDADAPNTPAWEAVYTL-NDGQGFVVTTN 424
YQ 383 PESNOGILTRKGLDFAKNOHTLYVEVNEAPFVLMKLPSTATIVVHVDNEAPVFP 442
DB 425 PVNNDGILTKAGLDFAKQYILHAVTVNVFVSVLTSTATVTVDLVDNEGPIFVP 484
YQ 443 PSKVVEQEGITPGEPCVYTAEDPK-ENQKISYRILRDPAGLWAMPDPSGQVAVGTL 501
DB 485 PEKRVESDFGVQELITSYTAQEPDTEQKITYRIWRDTRNWLINPTGAISRAEL 544
YQ 502 DREDEQFVRNNIYEVWVLANDNGSPPTGTGTLTLLIDVNDHGPVPEPRQITICNQSPV 561
DB 545 DREDFEHVKNSTYALIIATDNGSPVATGTGTLTLLSDVNDNAPIPEPTIFFCERNPK 604
YQ 562 RHLNITDKLSPTSPPFOQLTDDSDIYWTAEVNE-EGDTVLSLKKFLKQDITYDVHLS 620
DB 605 PQVINIHADLPNTSPFFAELTHGRVPNNTIQYNDPTQESILLKPKMALEVDGYKINLK 664
YQ 621 LSHGNEQLTVIRAVCDCHGVETC--PGPWKGF----ILPVGLAVALLFLLVLL 674
DB 665 LMDNQNKQDVTLEVSVCDEGAAGVCRKAQPVZAGLQIPAILGILGLALLILLLL 724
YQ 675 LTVKXKKEPILLPDDTRDNVYVYEGEGGEEDDYDTQLHGLEARPEVVLNDV 734
DB 725 LFLRRRAVKEPLPDDTRDNVYVYDEGGEEDQDFDLQHLHGLDAREVT-RNDV 783
YQ 735 APTIIPMYRPRANPEIGNFIENLKAANTDPTAPPYDTLLVPDYEGSGSDAASLSS 794
DB 784 APTLMSVRYLPRANPEIGNFIDENLKAADTPTAPPYDSILVFDYEGSGSEASLSS 843
YQ 795 LTSASDQDQDYDYLNEWSRPFKLADMYGGGED 828
DB 844 LNSSESDQDQDYDYLNEWGNRFKLLADMYGGGED 877

RESULT 9

US-09-798-267-2
Sequence 2, Application US/09798267
Patent No. 6406870

GENERAL INFORMATION:
APPLICANT: Brenner, Michael
APPLICANT: Cepek, Karyn

TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin
FILE REFERENCE: L0560/7008RP
CURRENT APPLICATION NUMBER: US/09/798,267
CURRENT FILING DATE: 2001-03-02

PRIOR APPLICATION NUMBER: US 08/237,919
PRIOR FILING DATE: 1994-05-03

PRIOR APPLICATION NUMBER: PCT/US 95/05518
PRIOR FILING DATE: 1995-05-03

PRIOR APPLICATION NUMBER: US 08/732,429
PRIOR FILING DATE: 1996-11-01

NUMBER OF SEQ ID NOS: 9

SOFTWARE: Patentin version 3.0

SEQ ID NO 2

LENGTH: 878

TYPE: PRT

ORGANISM: Homo sapiens

US-09-798-267-2

Query Match 53.8%; Score 2348.5; DB 4; Length 878;
Best Local Similarity 54.0%; Pred. No. 1.4e-209;
Matches 472; Conservative 115; Mismatches 230; Indels 57; Gaps 12;

YQ 8 LASLLQVLCWCAASEPCRAVFEAEVTLFAGGAEQEPQALGKV-FMGCPCGQ-----62
DB 8 LSGLLLLRSPLGQERSPPCLTRFLHVGAPAPPEKRP--RLGRVNFEDCTGQRTAI 65
YQ 63 -----PALFSTDNDDFTVRNGETVQERRSLK-----ERNP-----92
DB 66 FLTPIPKVGTDGVTIKRPLRFHNPDPFLGLRWDSYRKFSKVTLTNTVGHHRPPPHQ 125
YQ 93 -----LKIIP-SKRIIRHRKRDWVAPISVPENGKGFPPQRLNOLKSKNDRTKIF 142
DB 126 ASVSGIQAELLTFPNSPGLRRQRDWIPIPCPENEGKGFPPKNLVQKSKNKEGKVF 185
YQ 143 YSITGPGADSPPEGVFAVEKETGMLLNKPLDREEIAKYELFCHAVSENGASVEDPMNIS 202
DB 186 YSITGQADTPPVGVFIERTGMLKVTPELDREARIATYTLFHAVSSNGNAVEDPWEIL 245
YQ 203 IIVTDQNDHKPKTQDTPFGSVLEGVLPGTSVMQVATDEDDAIYTYNGVVAISHQEP 262
DB 246 IIVTDQNDKPEFTQEVFKGSVNEGALPGTSVMEVATDADDVNTYNAAIATYILSDP 305
YQ 263 KDPHLMFTIHRSTGTSVISGLDREKVPETLTQATDMGDGSGTTTAVAVVEILDAN 322
DB 306 ELDPKWMFTINRNTGVISVVTGLDRESFPTTLVQVQADLQGEGLSTATAVITVTDN 365
YQ 323 DNAPMFDPOKYEAHVPENAVCHVEQRLTVTDLPNPSPAWRATYILMGDDGDGHTTTH 382
DB 366 DNPPIFNPTTKYQVPEANVITLKVTDADAPNTPAWEAVYTL-NDGQGFVVTTN 424
YQ 383 PESNOGILTRKGLDFAKNOHTLYVEVNEAPFVLMKLPSTATIVVHVDNEAPVFP 442
DB 425 PVNNDGILTKAGLDFAKQYILHAVTVNVFVSVLTSTATVTVDLVDNEGPIFVP 484
YQ 443 PSKVVEQEGITPGEPCVYTAEDPK-ENQKISYRILRDPAGLWAMPDPSGQVAVGTL 501
DB 485 PEKRVESDFGVQELITSYTAQEPDTEQKITYRIWRDTRNWLINPTGAISRAEL 544
YQ 502 DREDEQFVRNNIYEVWVLANDNGSPPTGTGTLTLLIDVNDHGPVPEPRQITICNQSPV 561
DB 545 DREDFEHVKNSTYALIIATDNGSPVATGTGTLTLLSDVNDNAPIPEPTIFFCERNPK 604
YQ 562 RHLNITDKLSPTSPPFOQLTDDSDIYWTAEVNE-EGDTVLSLKKFLKQDITYDVHLS 620

Db 605 PQVINIHADLPNTSPFTAEHLTHGRVFNWNTIQNDPTQESIIKPKMALEVDYKINLK 664
 Qy 621 LSDHGNKEQTVIRATVCDCHGVETC--PGPWKGGF----ILPVLGAVLALLFLVL 674
 Db 665 LMDNQKQDQVTTLEVSVCDEGAAGVCRKAQPVAGLQIPAILGILGGLALLLILLLL 724
 Qy 675 LLVRKKKIKPELLPEDDTRDNVYVYEGGGGSEDDQYDITOLHRLGLARBPWLRLNDV 734
 Db 725 LFLRRRAVVEPLLPEDDTRDNVYVYDEEGGEEBDDFLSQHRLGLARPEVT-RNDV 783
 Qy 735 APTIIPMYRPRANPDEIGNFIENLKAANTDPTAPPYDITLLVFDYEGSGDAASLSS 794
 Db 784 APTLMSVPRYLPRANPDEIGNFIDENLKAADTPTAPPYDLSLLVFDYEGSGEASLSS 843
 Qy 795 LTSASDQDQDYLYNEWGSRFKKLADMYGGGD 828
 Db 844 LNSESDKQDQDYLYNEWGSRFKKLADMYGGGD 877

RESULT 10
 US-09-798-267-3
 ; Sequence 3, Application US/09798267
 ; Patent No. 6406870
 ; GENERAL INFORMATION:
 ; APPLICANT: Brenner, Michael
 ; APPLICANT: Cepek, Karen
 ; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin
 ; TITLE OF INVENTION: Interactions with T Lymphocytes
 ; FILE REFERENCE: L0560/7008ERP
 ; CURRENT APPLICATION NUMBER: US/09/798,267
 ; CURRENT FILING DATE: 2001-03-02
 ; PRIOR APPLICATION NUMBER: US 08/237,919
 ; PRIOR FILING DATE: 1994-05-03
 ; PRIOR APPLICATION NUMBER: PCT/US 95/05518
 ; PRIOR FILING DATE: 1995-05-03
 ; PRIOR APPLICATION NUMBER: US 08/732,429
 ; PRIOR FILING DATE: 1996-11-01
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: PatentIn.version 3.0
 ; SEQ ID NO 3
 ; LENGTH: 878
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-798-267-3

Query Match 53.8%; Score 2348.5; DB 4; Length 878;
 Best Local Similarity 54.0%; Pred.No.1.4e-209;
 Matches 472; Conservative 115; Mismatches 230; Indels 57; Gaps 12;
 Qy 8 LASLILLOVQWLCQAAEPCEAVREAEVLEAGGAEQPCQALGV-FMGCPQGE----- 62
 Db 8 LSGLLLLRLSPGLSQERSPPCLTRELHVGAPAPPEKRP--RLGRVNFEDCTGRTAI 65
 Qy 63 -----PALFSTDDNDFTRNGETVQERRSLK-----ERNP--- 92
 Db 66 FLTPPKVGTGDTVTKLEPLRFNHTDPLFLRWDSYRKSTKVLNTVGHRRPPHQ 125
 Qy 93 -----LKIFF-SKRILRRHKGDWVAPISVPENGKPPPPORLNQKSNKORDTKIF 142
 Db 126 ASVSGIOAELLTFPNSPGLRRQRDWDVIPPISCPENEGKPPFPKLVQIKSNKDKGKVF 185
 Qy 143 YSITPGGADSPGEGVAVEKETGMLLNKPLDRERIAKYELFHAVENGASVEDPMNIS 202
 Db 186 YSITGQGGADTPVGVFIERTIGLWTEPLDRERIAKYELFHAVENGASVEDPMNIS 245
 Qy 203 ITVDQNDHKPKFTQDTFRGVLSEGLVPGTSMVQVMTATDEDDAIYTVNGVVAISIHQEP 262
 Db 246 ITVDQNDKPEFTQEVFKGSMEGALPGTSMVMTATDADDDVNTYNAAIATYLSQDP 305
 Qy 263 KDPHDLMTIHRSGTISVSGGLDRKVPYTLTIQATMDGSGSTTTAVAVVEILDAN 322
 Db 306 ELPDKNMTINRNTGVISVVTIGLDRESFPYITLVQQAADLQEGSLTATAVITVTDN 365

RESULT 11
 PCI-US95-05518-2
 ; Sequence 2, Application PC/TUS9505518
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic
 ; TITLE OF INVENTION: E-cadherin Interactions with T Lymphocytes
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Wolf, Greenfield & Sacks P.C.
 ; STREET: 600 Atlantic Avenue
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/05518
 ; FILING DATE: herewith
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/237,919
 ; FILING DATE: 3 May 1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Plumer, Elizabeth R
 ; REGISTRATION NUMBER: 36,637
 ; REFERENCE/DOCKET NUMBER: B0801/7023
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-720-3500
 ; TELEFAX: 617-720-2441
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:

Qy 323 DNAPMFQKYEAHVPENAVGHVEQRLTVTDLDAPNSPAWRATYLMGDDGDGHFTTTH 382
 Db 366 DNPIFNPTTYKQVPEANVWITTLKVTDADAPNTPAWEAVVTL-NDGQGFVVTTN 424
 Qy 383 PESQGLITTRKGLDPRKQNTLYVETNEAPVFLKLPSTATIVVHVEDNEAPVFP 442
 Db 425 PVNNDGILTKAGLDFAKQYILHVAVTNVVPEVSLTSTATVTVVDVLDVNEGPVFP 484
 Qy 443 PSKVEVQEGIPTEGVCVTAEDPK-ENOKISYRILRDPAGWLAMDPSQGVTAAGTL 501
 Db 485 PEKRVESVEDGVCQETSITQAQPDTFMEQKITYRWRTRNWLINPTGALSTRAEL 544
 Qy 502 DREDEQFVRNNIYEVWLAMDNGSPPTGTGTLTLLTLDVNDHGPVPEPQITCNQSPV 561
 Db 545 DREDPEHVKNSTYATLIIATDNGSPVATGTGTLTLLSVDNDNAPIPEPRTIFCERNPK 604
 Qy 562 RHVLNITDKLSPHTSPQAOLTDSDIYTAENB-EGDTVVLSSKKFKLQDQTYDVHLS 620
 Db 605 PQVINIHADLPNTSPFTAEHLTHGRVFNWNTIQNDPTQESIIKPKMALEVDYKINLK 664
 Qy 621 LSDHGNKEQTVIRATVCDCHGVETC--PGPWKGGF----ILPVLGAVLALLFLVL 674
 Db 665 LMDNQKQDQVTTLEVSVCDEGAAGVCRKAQPVAGLQIPAILGILGGLALLLILLLL 724
 Qy 675 LLVRKKKIKPELLPEDDTRDNVYVYEGGGGSEDDQYDITOLHRLGLARBPWLRLNDV 734
 Db 725 LFLRRRAVVEPLLPEDDTRDNVYVYDEEGGEEBDDFLSQHRLGLARPEVT-RNDV 783
 Qy 735 APTIIPMYRPRANPDEIGNFIENLKAANTDPTAPPYDITLLVFDYEGSGDAASLSS 794
 Db 784 APTLMSVPRYLPRANPDEIGNFIDENLKAADTPTAPPYDLSLLVFDYEGSGEASLSS 843
 Qy 795 LTSASDQDQDYLYNEWGSRFKKLADMYGGGD 828
 Db 844 LNSESDKQDQDYLYNEWGSRFKKLADMYGGGD 877

; LENGTH: 878 amino acids	
; TYPE: amino acid	
; TOPOLOGY: linear	
; MOLECULE TYPE: protein	
PCT-US95-05518-2	
Query Match	
Best Local Similarity 53.8%; Score 2348.5; DB 5; Length 878;	
Matches 472; Conservative 115; Mismatches 230; Indels 57; Gaps 12;	
QY	8 LASLLQLQVWLCQAASEPCRAVFEAEVTLFAGGAEPCQALGV-FMGCPGQSE---62
DB	8 LSGLLLLRLPLSGQSESPPPCLTRFLVHGAPAPPEKRP-RLGRVNFEDCTGRQRTAI 65
QY	63 -----PALFSTDDNDFTVRNGETVQERRSLK-----ERNP---92
DB	66 FLTPRPKVGTDGVTIKRPLRFHNPDPFLGLRWSTYRKFSTKVTLNTVGHHRPPPHQ 125
QY	93 -----LKIRP-SKRILRRKRDWVAPISVPENGKGFPPQRLNOLKSNKDRDTKIF 142
DB	126 ASVSGIQAEILTPNPSFGRLRQKRDWIPPIPCPENEGKGFPPKRLVQIKSNKDKGKVF 185
QY	143 YSITPGADSPPEGVFAVEKETGMLLNKLPDREELAKYELFGHVSNGASVEDPMNIS 202
DB	186 YSITGQADTPPVGVFIERTGWLKVTEPLDRERIATYLFSSHVSSNGNAVEDPWEIL 245
QY	203 IIVTDQNDHKPKFTQDFRGSVLEGLVPGTSVMQVATDEDDAIYNGVVAISHQEP 262
DB	246 ITVTDQNDKPEFTQEVFKSGVMGALPGTSVMEVTAIDADDVNTYNAIAYTILSQDP 305
QY	263 KDPHDLFTIHRSTGTISVISGLDREKVPETILTQATMDGDGSTTTAVAVEILDAN 322
DB	306 ELPDKMFTINRNTGVISVTTGLDRESFPTYLIVQAAQLQEGSLSTATAVITVDIN 365
QY	323 DNAPMFDQKYEAHVPENAVGHEVQRLVTVDLDAAPSAPWARYLLIMGGDGDHFTITH 382
DB	366 DNPPFIFNTYKGVPEANEANVITLKVTDADAPNTPAWEAVYTIL-NDGCGQFVVTIN 424
QY	383 PESNGILLTRKGLDREKQHTLYEVETNEAPFVLKLPSTSTATIVVHVEDVNEAPVFP 442
DB	425 PVNNDGLKTKAGLDFEAKQYILHVAIVNVFVSEVSLTSTATIVTVLDVNEGFIVP 484
QY	443 PSKVEVEQEGIPTEPCVVTYABDDPK-ENQKISYILRDPAGWLAMPDPSGQVTAAGTL 501
DB	485 PEKRVESVEDFGVQSEITSYTAQEPDTFMEQKITRYIRWDRTRNLWLEINPDTGAISRAEL 544
QY	502 DREDEQFVRNNIYEVWVLMNDGSPPTGTGTLILLTLIDVNDKGPVPEPQITICNQSPV 561
DB	545 DREDFEHVKNSTYALIIATDNGSPVATGTGTLILLLSVDVNDNAPPEPRTIFFCERNPK 604
QY	562 RHVINITDKLSPTSPFQALTDSDIYVTAENVB-EGDTVVLKFLKQDTYDVHLS 620
DB	605 PQVINIHADLPNTSPFTAEHTHGRVPNNWTIQYNDPTQESILKPKMALEVGDYKINLK 664
QY	621 LSDHNGEQLTVIRATVCDCHGVETC--PGPWKGF-----ILPVLGAVALLFLLLVLL 674
DB	665 LMDNQNDQVTTLEVSVCDCGAGVCRKAQPVAGLIQIPAILGILGGILALLILILL 724
QY	675 LLVRKKEKIKPEILLPDDTRDNVYVYEGEGGEEQDYDITOLHGLEARPVWLRNDV 734
DB	725 LFLRRRAVKEPLLPPEDDTRDNVYVYDEEGGEEQDPLSLQHLGLDARPEVT-RNDV 783
QY	735 APTIIPPMYRPPANPDEIGNFIENLKAANTDPTAPPYDTLLVFDYEGSGDAASLSS 794
DB	784 APTLMSVRLPRPANPDEIGNFDENLKAADTDPTAPPYDLSLLVFDYEGSGEASLSS 843
QY	795 LTSASDQDDYDYLNEWGSREKFLADMYGGED 828
DB	844 LNSSESQDDYDYLNEWGNRFKFLADMYGGED 877
RESULT 12	
US-08-474-067-7	

; Sequence 7, Application US/08474067
; Patent No. 5811518
; GENERAL INFORMATION:
; APPLICANT: Rauscht, Barbara
; TITLE OF INVENTION: T-cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,067
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1682
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-474-067-7

Query Match	
Best Local Similarity 51.1%; Score 2231.5; DB 2; Length 837;	
Matches 450; Conservative 117; Mismatches 215; Indels 55; Gaps 13;	
QY	48 GOALGV-FMGCPGQEPALF-STD-----NDD-----FIVRNCETVQE 83
DB	6 GEELGRVSFAACSGRPWAVYVPTDTRFKVNGDGVVSTKRPLTLYGRKISFTIYQADAWGK 65
QY	84 RES-----LKERNPDKI-FP--SKRILRRKRDWVAPISVPENGK 122
DB	66 RLSARVTVGRHRRHRRHHNHLLQDTPAVLTPKXDPGLRQKEDWIPISCLNHRG 125
QY	123 PPQRLNQLKSKORDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKLPDRREIAKYE 182
DB	126 PYFMELVQIKSKDKESKVIYSITGQADSPPVGFIERTGMLVEVTEQDLREKIDRYT 185
QY	183 LFGHAVSNGASVEDPMNISIIIVTDQNDHKPKFTQDTERGSVLEGLVPGTSVMQVATDE 242
DB	186 LLSHAVSASGQVEDPMEIITVMDQNDKVPFKEVFGVIEENAKGCTSVMTVATDA 245
QY	243 DDATVYNGVWVAYSITHSQPKDHPDLMTIHRSTGTISVISGLDREKVPYTYLTIOATD 302
DB	246 DDVAVTNDGIVSYSVSQPPRHPQMTIDPAKGIISVLGTGLDRETPNTYTLIVQATD 305
QY	303 MDGDSSTTAVAVVIELDANDNAPMFDQKYEAHVPENAVGHEVQRLVTVDLDAENSPAW 362
DB	306 QEGKGLSNTATAILEVTDANDNIPFNFTWEGVVEENKPGTEVARLTVTQODAPGSPAW 365
QY	363 RATYILMGDDGDHFTITTHPESNGOILTRKGLDFAKNQHTLYEVETNEAPFVLKLP 422
DB	366 QAVYHIKSGNLDGAFSIITDPTNNGILKTKAGLDYETKSYDVLVVTVENKVLSPVITL 425

423 STATVHVHEDVNEAPVFPVPPSKVVEQVGPVGVVTAEDPKE-NOKISYRILRD 481
426 STASVLTVDVNEPPVFPVPIKRVGVPEPLVGVQVTSYTAEDPDRMRQKITRMSGD 485
482 PAGWLAMPDQSGOVAVGTLDREDEQVRNIVYEVWVLANDGSPPTGTGTLTLLIDV 541
486 PAGWLYIHPENGIVTATQPDRESVHAI-NSTYKAILAVDNGIPDTTGTGTLTLLIDV 544
542 NDHGFVPEPRQITTCNOSPVRHVLNITDKLSPHTSFQAOLTDSDIYTAEVNEEGDT 601
545 NDNGTPEPRSFECISQPEKQILSVDKLPHTYFKAALHGGSSNNVTVEIRQ-DE 603
602 VVLSLKFLKQDVTYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCP--GPWKGGF--- 655
604 LAMGLKKELEPGEYNI-FVKLTDSQKQAVTQVKAQVCEGTAKNCERRSYIVGGLGVA 663
656 ILPVLGAVLALLFLLVLLVLRKRIKEPPLLPEDDTRDNVFFYEGEGGEDDQYDI 715
664 ILGILGILALLIILLVLLFARRKVEKEPPLPPEDDMRDNVYDEEGGEDDQYDL 723
716 TOLHGLEARPEVLRNDVAPTIIPTMYRPRPANDEIGNFIENLKAANTDPTAPPYD 775
724 SQLHGLDARPEV-RNDVAPPLMAAPQYRPRPANDEIGNFIENLKAANTDPTAPPYD 782
776 TLLVFDYEGSGSDAASLSLTSASDQDQDYDYLNEWGSRFKKLADMYGGGEDD 829
783 SLLVFDYEGSGSEATSLSSLSNASDQDQDYDYLNEWGRFKKLAEYGGGEDD 836

RESULT 13

US-08-474-068A-7
; Sequence 7, Application US/08474068A
; Patent No. 5937525
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,068A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/213,361
; FILING DATE: 14-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/607,293
; FILING DATE: 30-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1683
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-474-068A-7

Query Match 51.1%; Score 2231.5; DB 2; Length 837;
Best Local Similarity 54.0%; Pred. No. 1-le-196;
Matches 450; Conservative 117; Mismatches 212; Indels 55; Gaps 13;
QY 48 GOALGKV-FMGCQGPALF-STD-----NDD-----FTVRNGETVQE 83
DB 6 GRELGRVSFACSGRPWAVYPTDTRFVNGDGVVSNKRPILTYGRKISFTIYAQDAMG 65
QY 84 RRS-----LKERNPULI-FP--SKRILRRHRKRDWVADISVPEKNGK 122
DB 66 RRSARVTVGRHRRHRRHHHLLQDTPFAVLTFPPKHDFGRQRKRDWVIPPISLEHNR 125
QY 123 PPORLNLKSKKORDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREREIAKYE 182
DB 126 PYPMLVQIKNKDKESKVYISITGGADSPVGVIFIERETGHWLEVEQDLDRKIDRYT 185
QY 183 LFGHAVSENGASVEDPMNISIIITDQNDHKPKFTQDTPRGSVLEGLVPGTSMVQVATDE 242
DB 186 LLSHAVSASGQVDEPMBIIITVMDQNDKPVFIKEVFVGIENAKPGTSMVTNATDA 245
QY 243 DDALVYNGVAVSYTHSQEPKDPHDLMTIHRSTGTSIVSSGLDRKVPETVLTIOATD 302
DB 246 DAVNTDNGIVSYISVSOQPPRPHQPTIDPAKGIISVLGTGLDRETTPNYTLIVQATD 305
QY 303 MDGSGTITTAVAVVEILDANDNAPMDFQKYEAHVPENAVAGHEVORLTVTDLDAPNSPAW 362
DB 306 QEGKLSNTATAIETVTDANDNIPFNPTMYEGVVEENKPGTEVARLTVTDQDAPGSPAW 365
QY 363 RATYLIMCGDDGHDFTITTHPESNOGILTTKGLDPEAKNOHTLYVEVNEAPVXLPT 422
DB 366 QAVYHIKSNLGDGAFSIITDESTNGILKTAGLDYETKSRDYLDVAVTVENKVPUSVITL 425
QY 423 STATIWHVEDVNEAPVFPVPPSKVVEQVGPVGVVTAEDPKE-NOKISYRILRD 481
DB 426 STASVLTVDVNEPPVFPVPIKRVGVPEPLVGVQVTSYTAEDPDRMRQKITRMSGD 485
QY 482 PAGWLAMPDQSGOVAVGTLDREDEQVRNIVYEVWVLANDGSPPTGTGTLTLLIDV 541
DB 486 PAGWLYIHPENGIVTATQPDRESVHAI-NSTYKAILAVDNGIPDTTGTGTLTLLIDV 544
QY 542 NDHGFVPEPRQITTCNOSPVRHVLNITDKLSPHTSFQAOLTDSDIYTAEVNEEGDT 601
DB 545 NDNGTPEPRSFECISQPEKQILSVDKLPHTYFKAALHGGSSNNVTVEIRQ-DE 603
QY 602 VVLSLKFLKQDVTYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCP--GPWKGGF--- 655
DB 604 LAMGLKKELEPGEYNI-FVKLTDSQKQAVTQVKAQVCEGTAKNCERRSYIVGGLGVA 663
QY 656 ILPVLGAVLALLFLLVLLVLRKRIKEPPLLPEDDTRDNVFFYEGEGGEDDQYDI 715
DB 664 ILGILGILALLIILLVLLFARRKVEKEPPLPPEDDMRDNVYDEEGGEDDQYDL 723
QY 716 TOLHGLEARPEVLRNDVAPTIIPTMYRPRPANDEIGNFIENLKAANTDPTAPPYD 775
DB 724 SQLHGLDARPEV-RNDVAPPLMAAPQYRPRPANDEIGNFIENLKAANTDPTAPPYD 782
QY 776 TLLVFDYEGSGSDAASLSLTSASDQDQDYDYLNEWGSRFKKLADMYGGGEDD 829
DB 783 SLLVFDYEGSGSEATSLSSLSNASDQDQDYDYLNEWGRFKKLAEYGGGEDD 836

RESULT 14

US-08-472-481-6
; Sequence 6, Application US/08472481
; Patent No. 5863804
; GENERAL INFORMATION:
; APPLICANT: Ranscht, Barbara
; TITLE OF INVENTION: T-Cadherin Adhesion Molecule
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,481
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/213,361
FILING DATE: 14-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/607,293
FILING DATE: 30-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1686
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8943
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 837 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-472-481-6

Query Match 51.1%; Score 2231.5; DB 2; Length 837;
Best Local Similarity 54.0%; Pred. No. 1.1e-198;
Matches 450; Conservative 117; Mismatches 212; Indels 55; Gaps 13;

QY 48 GOALGKV-FMGCPQOEALP-STD-----NDD-----PTVANGTVQS 83
DB 6 GRELGVSFAACGRPAWVTPDTRFKNGDGVSTYKRLTYLGRKISFTIYAQAMGK 65
QY 84 RRS-----LKERNFLKI-FP--SKRIILRRHKRDWVAPISVPENGK 122
DB 66 RHGARTVGRHRRHRRHHHLQDTPAVLTFPKHDPGLRRQKRDWVIPPISLENHRG 125
QY 123 PFQRLNQLSKDRDRTKIPYSITGQADSPPGVFAVETGWLKLLNKLDRREEAKYE 182
DB 126 FYPMLVQIKSNKDKESKVYSITGQADSPPGVFIETRETGLWLVTEQLDREKIDRYT 185
QY 183 LFHGAHSENGASVEDPMNIIIVTDQNDHKPKFTQDTRFGSVLEGVLPGTSVMQVTAIDE 242
DB 186 LLSHVSASQPVEDPMELIITVMDQNDKKEVFIKEVFGYIEENAKPGTSVMVATDA 245
QY 243 DDALYTVNGVAVSIHSGEPKPDHLMFTTHRSTGTISVISGLDREKVPYTLTIQATD 302
DB 246 DDAVNTONGVSVSIVSQPPRPQFTIDPAKGIISVLGTGLDRETTENYTLIVQATD 305
QY 303 MDGSGSTTAVAVVEILDANDNAPMFPQKYEAPVAVENAVGHEVQRLTVTDLPAPNSPAW 362
DB 306 QEGKGLNTATAILVTDANDNPIFNPTMYEGVVENKEKTEVARELTVTDQAPGSPAW 365
QY 363 RAYLYMGDDGDHFTITTHPESNGGLITRKGDLDFEAKNQHTLYVEVNEAPFVLKLP 422
DB 366 QAVYHIKSGNLGDAFSIITDPSTNNGILKTAAGLDYETKSRDYLVTVVENKVPISVITL 425
QY 423 STAVIVHVEDVNEAPVFPVPSKVVEQGIPTGEPVCVTAEDPQKE-NOKISYRLRD 481
DB 426 STASVLTVLDVNEPFPVFPVPPKVGVPEDLPVGCQVTSYTABDPDRMKQITYRMGSD 485
QY 482 PAGMLAMPDPSGVQVAVGTGLDREDEQFVRNNIYVEMVLMADNGSPPTTGTGTLILLIDV 541
DB 486 PAGWLYIHPENGIIVTATQPLDRESVHAI-NSTYKAILIAYDNGIPDTGTGTILLILLQDV 544

QY 542 NDHGFVPEPRQITICNQSPVRHVLNITKDLSPHTSPFOAQLTDDSDIYWTAEVNEBGT 601
DB 545 NDNGPTPEPRSGFEICSRQPEKQILSIVDKLPHTYTPFKAALEGSSNNWTVIRGQ-DE 603
QY 602 VVLSLKKEFKODTYDVHLSDHGNKEQLTVIRATVCDCHGHVETCP--GPWKGF----- 655
DB 604 LANGLKSELEPGYNIFFVKLTDSQGAQVTVKQAQVCEGTAKNCERRSRIVVGLGVA 663
QY 656 ILPVLGAVLALIFLLLVLLLVLRKKRKEPLLLPEDDTRDNVFFYEGEGGEEEDQDYDI 715
DB 664 ILGILGGILALLLLLLLLFARRRKEVEPLLPEDDVRNVVNYDEEGGEEEDQDYDL 723
QY 716 TOLHRLGARBEVVLNDVAPTIITPMYRPRPANDPEIGNFIENLKAANTDPTAPYD 775
DB 724 SQLHGLDARPEVI-RNDVAPPLMAAPQYRPRPANDPEIGNFIENLKAANTDPTAPYD 782
QY 776 TLLVEDYEGSGDAASLSLTSASDQDQDYDLNEMGSRFKKLADMTYGGGDD 829
DB 783 SILVFDYEGSGSEATSLSSLSASDQDQDYDLNEMGNRRFKLAELXGGGDD 836

RESULT 15
US-09-417-039-11
Sequence 11, Application US/09417039A
Patent No. 6485972
GENERAL INFORMATION:
APPLICANT: McMahon, Andrew P.
APPLICANT: Parr, Brian A.
APPLICANT: Vaino, Seppo
TITLE OF INVENTION: WNT SIGNALING IN REPRODUCTIVE ORGANS
FILE REFERENCE: 00246/232001
CURRENT APPLICATION NUMBER: US/09/417,039A
CURRENT FILING DATE: 1999-10-12
EARLIER APPLICATION NUMBER: US 60/109,355
EARLIER FILING DATE: 1998-10-15
NUMBER OF SEQ ID NOS: 11
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11
LENGTH: 906
TYPE: PRT
ORGANISM: Homo sapiens
US-09-417-039-11

Query Match 39.2%; Score 1714; DB 4; Length 906;
Best Local Similarity 42.8%; Pred. No. 2.3e-150;
Matches 391; Conservative 129; Mismatches 291; Indels 102; Gaps 27;

QY 6 GLASLILQLVCWLQCAASEP-----CRAVPREAVTLEAGABOEPOALGKV-FMGCP 59
DB 6 GALRTLLPILLALLQASVEASGEIALCKTGPE-DVYSAVLSKDVHEGQPLLNVKFSNCN 64
QY 60 GO-----EPALFSTND--FTVRN-----GETVOER----- 84
DB 65 GRRKQVRESSEPADPKVEDGMVAVRSPFLSSSEHAKFLIYAQDKETQEKQVAVKLSL 124
QY 85 -----RSLKERNPLK--IFP---SKRI--LRRHKRDWVAFISVPENKGPFPQRLNQL 132
DB 125 PTLTEESVKESAEEVIFVPRQFSKHGLOKQKEDWIPPIINLPENSRRGPPPOELVRIR 184
QY 133 SNKDRDRTKIFYITGCGADSPGEGVFAVEKSTGWLKLLNKLDRREIAKVELFGHVSNG 192
DB 185 SDRDKNLSLRYSVTGPGADQPTGIFINPISGQSVTKPLDREQIARFHLRAHVAIDNG 244
QY 193 ASVEDPMNIIIVTDQNDHKPKFTQDTRFGSVLEGVLPGTSVMQVTADEDDAIYTVNGV 252
DB 245 NQVENPIDIVINVDNDRNRPFLHQVNGTVPEGSKGTYVMTVTAIDADDP-NALNGM 303
QY 253 VAYSHSGEPKPDHLMFTTHRSTGTISVISGLDREKVPYTLTIQATDGDG---GST 309
DB 304 LRYRIVSQAPSPSPNMTINNETGDIITVAAGLDREKVVQYTLTIQATDMEGNTYGLS 363
QY 310 TTAVAVVEILDANDNAPMFPDQKYEAPVAVENAVGHEVQRLTVTDLPAPNSPAWRATYILM 369

Db 364 NTATAVITVTDVNDNPEFTAMTYGEVPEPNRVDIIVANLTVTDKQOPHTPAMNAVYRIS 423
Qy 370 GGDGDHFTIITHPESNOGLTTRKGLDFAKQOHTLYVEVTNEAPEVLKL---PTSTAT 426
Db 424 GGDPTGRFAIQTDENSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQHPPQSTAT 483
Qy 427 IVVVEDVNEAPVFPVPSKVVEOEGIPTGEPVCVYTAEDBDK-ENQKISYRILRDPAGW 485
Db 484 VSVTVIDVNEPNFYAPNPKIIRQEBGLHAGTMLTFTTAQDPDRYMQONIRYTKLSDPANW 543
Qy 486 LAMPDSCQVAVGTLDREDEQFVRNNIYEVWVLAMDNQSPPTGTGTLTLLTLDVNDHG 545
Db 544 LKIDPVNGQITTIIVLDRESEN-VKNNIYNATFLASDNQIPPMMSGTGTLOIYLLDINDNA 602
Qy 546 PVPEPRQITICNQSPVRHVLNIT--DKDLSPHTSPPFOAQLTDDSDIY---WTAEVNEEGD 600
Db 603 PQVLFQEAETC-ETFPDPSINITALDYDIDPNAGFFAFDLSPVTIKRNWTI-TRLNGD 660
Qy 601 TVVLSLK-KFLKQDTYDVHLSLSDHGN--KEQLTVIRATVCDCHG-----VETCPGPWK 652
Db 661 FAQLNLKIKFLEAGIYEVPIIITDSGNPPKSNISILRVKVCQDSNGDCTDVRIVGAGL 720
Qy 653 GGFILPVLGAVLALLFLLVLLLV-----RKGRKIKPELLLPEDDTRDNVFFYG 702
Db 721 G-----TGATIAILLCTIIILLILVLMFVVMKRRDKERQAKQLLIDPEDDVRDNILKYD 774
Qy 703 BEGGGEEDDYDITQLHGLEAREPV-----VLENDVAPTIIPTMYRPERA--NPDEIG 755
Db 775 BEGGGEEDQDYDLSQLQPDPTVEPDAIKPVGIRMDERP-IHAEPQYFVRSAAHPHGDIG 833
Qy 756 NFIIENLKAANTDPTAPPYDTLLVFDYEGSGSDAASLSLTSSASDQDQDYDYLNEWGSR 815
Db 834 DFINEGLKAADNDPTAPPYDLSLVFDYEGSGSTAGLSLSLSSSSSGGGEQDYDYLNDWGPR 893
Qy 816 FKKLADMYGGGED 828
Db 894 FKKLADMYGGGDD 906

Search completed: September 21, 2004, 22:10:10
Job time : 36 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:09:14 ; Search time 143 Seconds
(without alignments)
1861.675 Million cell updates/sec

Title: US-09-916-849A-1
Perfect score: 4369
Sequence: 1 MGLPRGLASLLLVQVWLQ.....NEWGSRFKLADMYGGEDD 829

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1342398 seqs, 321133274 residues

Total number of hits satisfying chosen parameters: 1342398

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	4369	100.0	829	9	US-09-905-983-50
2	4369	100.0	829	10	US-09-916-849A-1
3	4369	100.0	829	10	US-09-975-723A-1
4	4369	100.0	829	12	US-10-058-270A-126
5	4369	100.0	829	14	US-10-174-677-21
6	4369	100.0	829	14	US-10-158-123-1
7	4369	100.0	829	15	US-10-295-027-783
8	4369	100.0	829	15	US-10-295-027-800
9	4369	100.0	829	15	US-10-295-027-863
10	4369	100.0	829	15	US-10-295-027-896
11	4361	99.8	829	12	US-10-329-345-16
12	4361	99.8	829	12	US-10-274-177-16
13	4361	99.8	829	14	US-10-301-822-18
14	4361	99.8	829	15	US-10-295-027-338
15	4361	99.8	829	15	US-10-295-027-1246

16	4361	99.8	829	15	US-10-087-080-14	Sequence 14, Appl
17	4361	99.8	829	16	US-10-650-112-16	Sequence 16, Appl
18	2420	55.4	882	14	US-10-097-340-37	Sequence 37, Appl
19	2388	54.7	899	9	US-09-905-983-5	Sequence 5, Appl
20	2348.5	53.8	878	13	US-10-165-049-2	Sequence 2, Appl
21	2348.5	53.8	878	13	US-10-165-049-3	Sequence 3, Appl
22	2341.5	53.6	878	9	US-09-905-983-48	Sequence 48, Appl
23	2341.5	53.6	878	15	US-10-173-551-34	Sequence 34, Appl
24	2285	52.3	894	12	US-10-425-114-39106	Sequence 39106, A
25	2216.5	50.7	821	14	US-10-174-677-28	Sequence 28, Appl
26	1714	39.2	906	14	US-10-174-677-19	Sequence 19, Appl
27	1707	39.1	906	9	US-09-905-983-46	Sequence 46, Appl
28	1707	39.1	906	14	US-10-177-293-43	Sequence 43, Appl
29	1707	39.1	906	16	US-10-778-146-6	Sequence 6, Appl
30	1704.5	39.0	912	9	US-09-905-983-2	Sequence 2, Appl
31	1702	39.0	906	9	US-09-746-491-46	Sequence 46, Appl
32	1692.5	38.7	912	9	US-09-746-491-49	Sequence 49, Appl
33	1685.5	38.6	906	9	US-09-746-491-48	Sequence 48, Appl
34	1679	38.4	877	9	US-09-746-491-47	Sequence 47, Appl
35	1648.5	37.7	916	15	US-10-295-027-286	Sequence 286, App
36	1638	37.5	824	15	US-10-104-047-2515	Sequence 2515, Ap
37	1621.5	37.1	916	14	US-10-174-677-22	Sequence 22, Appl
38	1621.5	37.1	916	15	US-10-116-275-201	Sequence 201, App
39	1422	32.5	848	12	US-09-805-020-56	Sequence 56, Appl
40	1321.5	30.2	814	14	US-10-174-677-14	Sequence 14, Appl
41	1242	28.4	819	15	US-10-161-493-62	Sequence 62, Appl
42	1164	26.6	556	10	US-09-880-573-98	Sequence 98, Appl
43	1030.5	23.6	785	14	US-10-174-677-25	Sequence 25, Appl
44	1021.5	23.4	713	14	US-10-174-677-13	Sequence 13, Appl
45	1021.5	23.4	713	15	US-10-408-765A-394	Sequence 394, App

ALIGNMENTS

RESULT 1

US-09-905-983-50
; Sequence 50, Application US/0905983
; Patent No. US20020045591A1
; GENERAL INFORMATION:
; APPLICANT: Geiger, Benjamin
; APPLICANT: Ben-Ze'ev, Avri
; APPLICANT: Sadot, Einat
; TITLE OF INVENTION: METHODS AND THERAPEUTIC COMPOSITIONS FOR TREATING CANCER
; FILE REFERENCE: 01/22326
; CURRENT APPLICATION NUMBER: US/09/905,983
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-983-50

Query Match 100.0%; Score 4369; DB 9; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MGLPRGLASLLLVQVWLQCAASEPCRAVPREAVTLEAGGAEQPCGALGKVFMGCPG	60
Db	1	MGLPRGLASLLLVQVWLQCAASEPCRAVPREAVTLEAGGAEQPCGALGKVFMGCPG	60
QY	61	QEPALFSTNDQFTVRNGETVOERSLKERNPLKIPPSKRIILRRHKRDVWVAPISVPENG	120
Db	61	QEPALFSTNDQFTVRNGETVOERSLKERNPLKIPPSKRIILRRHKRDVWVAPISVPENG	120
QY	121	KGPFPPQRLNQLSKNDRDTKIFYSITGPGADSPPEGVFAVEKETGWLINLKPDLDEEIAK	180
Db	121	KGPFPPQRLNQLSKNDRDTKIFYSITGPGADSPPEGVFAVEKETGWLINLKPDLDEEIAK	180
QY	181	YELPHGVHSENGASVEDPMNIIIVTDONDHPKPTQDTFRGSLVLEGLVPGVSNVQVTAT	240

```

Db 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240
QY 241 DEDDAIYTYNGVAVSIHQSQPKDPHDLMTIHRSTGTISVISSGLDREKVPYTLTIOA 300
Db 241 DEDDAIYTYNGVAVSIHQSQPKDPHDLMTIHRSTGTISVISSGLDREKVPYTLTIOA 300
QY 301 TMDGCGSTTTAVAVEILDANDNAPMDPKQYEAHVPENAVHGVORLTVTDLDAPNSP 360
Db 301 TMDGCGSTTTAVAVEILDANDNAPMDPKQYEAHVPENAVHGVORLTVTDLDAPNSP 360
QY 361 AWRATYLMGDDGDGHFTIITHPESNOGILTRKGLDPEAKNOHTLVVETNEAPFVLK 420
Db 361 AWRATYLMGDDGDGHFTIITHPESNOGILTRKGLDPEAKNOHTLVVETNEAPFVLK 420
QY 421 PTSTATIIVHVEDVNEAPFVPPSKVVEVQSGIPTGEPVCVYTAEDPDKENQKISYRILR 480
Db 421 PTSTATIIVHVEDVNEAPFVPPSKVVEVQSGIPTGEPVCVYTAEDPDKENQKISYRILR 480
QY 481 DPAGWLAMPDSGGQVAVGTILDRDEQFVRNNIYEVWVLAMNDGSPPTTGTGLLLTLID 540
Db 481 DPAGWLAMPDSGGQVAVGTILDRDEQFVRNNIYEVWVLAMNDGSPPTTGTGLLLTLID 540
QY 541 VNDHGPVPEPQITICNOSPVRHVLNITDKLSPTSPPFOAQLTDDSDIYTAEVNEEGD 600
Db 541 VNDHGPVPEPQITICNOSPVRHVLNITDKLSPTSPPFOAQLTDDSDIYTAEVNEEGD 600
QY 601 TVVLSLKKFLKQDQTYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL 660
Db 601 TVVLSLKKFLKQDQTYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL 660
QY 661 GAVLALLFLLVLLVLLVLRKKIKKEPILLPDDTRDNVFFYGGEGEEDQDIDITQLHR 720
Db 661 GAVLALLFLLVLLVLLVLRKKIKKEPILLPDDTRDNVFFYGGEGEEDQDIDITQLHR 720
QY 721 GLEARPEVVLNDVAPTIPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
Db 721 GLEARPEVVLNDVAPTIPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
QY 781 DYEGSGDAASLSLTSSASDQDQDYDLNWSRFRKFLADMYGGEDD 829
Db 781 DYEGSGDAASLSLTSSASDQDQDYDLNWSRFRKFLADMYGGEDD 829

RESULT 2
US-09-916-849A-1
; Sequence 1, Application US/09916849A
; Publication No. US20030086934A1
; GENERAL INFORMATION:
; APPLICANT: Bostein, et al.
; TITLE OF INVENTION: Basal Markers in Breast Cancer and Related Reagents
; FILE REFERENCE: 2002850-0024
; CURRENT APPLICATION NUMBER: US/09/916,849A
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 829
; TYPE: PR1
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence of
; Cadherin 3
US-09-916-849A-1

Query Match 100.0%; Score 4369; DB 10; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGLASLLLLQVCWLQCAASEPCRAVFAEVTLEAGGAEQBPFGALGVFMGCPG 60
Db 1 MGLPRGLASLLLLQVCWLQCAASEPCRAVFAEVTLEAGGAEQBPFGALGVFMGCPG 60

```

```

QY 61 QBPALFSTNDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAPISVPENG 120
Db 61 QBPALFSTNDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAPISVPENG 120
QY 121 KGPFFORLNQLSKNDROTKIPIYSITGPGADSPPEGVFAVEKETGWLKLLNKLPLDREETA 180
Db 121 KGPFFORLNQLSKNDROTKIPIYSITGPGADSPPEGVFAVEKETGWLKLLNKLPLDREETA 180
QY 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240
Db 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240
QY 241 DEDDAIYTYNGVAVSIHQSQPKDPHDLMTIHRSTGTISVISSGLDREKVPYTLTIOA 300
Db 241 DEDDAIYTYNGVAVSIHQSQPKDPHDLMTIHRSTGTISVISSGLDREKVPYTLTIOA 300
QY 301 TMDGCGSTTTAVAVEILDANDNAPMDPKQYEAHVPENAVHGVORLTVTDLDAPNSP 360
Db 301 TMDGCGSTTTAVAVEILDANDNAPMDPKQYEAHVPENAVHGVORLTVTDLDAPNSP 360
QY 361 AWRATYLMGDDGDGHFTIITHPESNOGILTRKGLDPEAKNOHTLVVETNEAPFVLK 420
Db 361 AWRATYLMGDDGDGHFTIITHPESNOGILTRKGLDPEAKNOHTLVVETNEAPFVLK 420
QY 421 PTSTATIIVHVEDVNEAPFVPPSKVVEVQSGIPTGEPVCVYTAEDPDKENQKISYRILR 480
Db 421 PTSTATIIVHVEDVNEAPFVPPSKVVEVQSGIPTGEPVCVYTAEDPDKENQKISYRILR 480
QY 481 DPAGWLAMPDSGGQVAVGTILDRDEQFVRNNIYEVWVLAMNDGSPPTTGTGLLLTLID 540
Db 481 DPAGWLAMPDSGGQVAVGTILDRDEQFVRNNIYEVWVLAMNDGSPPTTGTGLLLTLID 540
QY 541 VNDHGPVPEPQITICNOSPVRHVLNITDKLSPTSPPFOAQLTDDSDIYTAEVNEEGD 600
Db 541 VNDHGPVPEPQITICNOSPVRHVLNITDKLSPTSPPFOAQLTDDSDIYTAEVNEEGD 600
QY 601 TVVLSLKKFLKQDQTYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL 660
Db 601 TVVLSLKKFLKQDQTYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL 660
QY 661 GAVLALLFLLVLLVLLVLRKKIKKEPILLPDDTRDNVFFYGGEGEEDQDIDITQLHR 720
Db 661 GAVLALLFLLVLLVLLVLRKKIKKEPILLPDDTRDNVFFYGGEGEEDQDIDITQLHR 720
QY 721 GLEARPEVVLNDVAPTIPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
Db 721 GLEARPEVVLNDVAPTIPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
QY 781 DYEGSGDAASLSLTSSASDQDQDYDLNWSRFRKFLADMYGGEDD 829
Db 781 DYEGSGDAASLSLTSSASDQDQDYDLNWSRFRKFLADMYGGEDD 829

RESULT 3
US-09-916-849A-1
; Sequence 1, Application US/09916849A
; Publication No. US20030108529A1
; GENERAL INFORMATION:
; APPLICANT: Focy, Ramsey
; APPLICANT: Nackman, Gary
; TITLE OF INVENTION: Improvement of Endothelial Cell-Cell
; FILE REFERENCE: 601-1-101N
; CURRENT APPLICATION NUMBER: US/09/916,849A
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 60/241,216
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/243,693
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 829

```

TYPE: PRT
ORGANISM: Homo sapiens
US-09-975-723A-1

Query Match 100.0%; Score 4369; DB 10; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGLASLLQLQVLCWLOCAASPCRAVFEAEVLEAGGAEOBPGQALGVFGCPG 60
DB 1 MGLPRGLASLLQLQVLCWLOCAASPCRAVFEAEVLEAGGAEOBPGQALGVFGCPG 60

QY 61 QEPALFSTDDNDFTVRNGETVQERRSLKERNPLKIPPSKRILRRHKRDWVAPISVPENG 120
DB 61 QEPALFSTDDNDFTVRNGETVQERRSLKERNPLKIPPSKRILRRHKRDWVAPISVPENG 120

QY 121 KGPFPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLNLKPLDRBEIAK 180
DB 121 KGPFPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLNLKPLDRBEIAK 180

QY 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSLVLEGLPGTSVMQVAT 240
DB 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSLVLEGLPGTSVMQVAT 240

QY 241 DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTIHRSTGTISVSSGLDREKVPYTLTIOA 300
DB 241 DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTIHRSTGTISVSSGLDREKVPYTLTIOA 300

QY 301 TDMGDGSTTTAVAVVEILDANDNAPMFPDQKYEAHVPENAVGEVORLTVTDLAPNSP 360
DB 301 TDMGDGSTTTAVAVVEILDANDNAPMFPDQKYEAHVPENAVGEVORLTVTDLAPNSP 360

QY 361 AWRATYIMGGDDGDHFTTIHPESNQILTRKGLDPEAKNQHTLYVEVNEAPFVLK 420
DB 361 AWRATYIMGGDDGDHFTTIHPESNQILTRKGLDPEAKNQHTLYVEVNEAPFVLK 420

QY 421 PTSTATIIVHVEDVNEAPVFPVPSKVVEQEGIPTEGVCVYTAEDDPKENOKISYRILR 480
DB 421 PTSTATIIVHVEDVNEAPVFPVPSKVVEQEGIPTEGVCVYTAEDDPKENOKISYRILR 480

QY 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMNDGSPPTTGTGLTLTLID 540
DB 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMNDGSPPTTGTGLTLTLID 540

QY 541 VNDHGPVPEPRQITICNQSPVRHLNITDKDLSPTSPFQAQLTDDSDIYWTAEVNEEGD 600
DB 541 VNDHGPVPEPRQITICNQSPVRHLNITDKDLSPTSPFQAQLTDDSDIYWTAEVNEEGD 600

QY 601 TVVLSLKKFLKQDVTYDVHLSLSDHGKNEQLTVIRATVCDCHGHVETCPGPKWGGFILPVL 660
DB 601 TVVLSLKKFLKQDVTYDVHLSLSDHGKNEQLTVIRATVCDCHGHVETCPGPKWGGFILPVL 660

QY 661 GAVLALLFLLVLLVLLVLRKKIKKEPILLPDDTRDNVFFYGGEGGEEDQDITOLHR 720
DB 661 GAVLALLFLLVLLVLLVLRKKIKKEPILLPDDTRDNVFFYGGEGGEEDQDITOLHR 720

QY 721 GLEARPEVVLNDVAPTIIPPMVRPRPANDENFTIENLKAANDTAPAPPVDTLLVF 780
DB 721 GLEARPEVVLNDVAPTIIPPMVRPRPANDENFTIENLKAANDTAPAPPVDTLLVF 780

QY 781 DYEGSGSDAASLSLSTSSASQDQDYDYLNEWGSFRFKKLADMYGGGDD 829
DB 781 DYEGSGSDAASLSLSTSSASQDQDYDYLNEWGSFRFKKLADMYGGGDD 829

RESULT 4

US-10-058-270A-126
Sequence 126, Application US/10058270A
Publication No. US20040029114A1
GENERAL INFORMATION:
APPLICANT: Mack, David H.
Gish, Kurt C.
Afar, Daniel

APPLICANT: Eos Biotechnology, Inc.
TITLE OF INVENTION: Methods of Diagnosis of Breast Cancer, Compositions and
TITLE OF INVENTION: Methods of Screening for Modulators of Breast Cancer
FILE REFERENCE: 018501-005210US
CURRENT APPLICATION NUMBER: US/10/058,270A
CURRENT FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: US 60/263,965
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: US 60/265,928
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: US 09/829,472
PRIOR FILING DATE: 2001-04-09
PRIOR APPLICATION NUMBER: US 60/282,698
PRIOR FILING DATE: 2001-04-09
PRIOR APPLICATION NUMBER: US 60/288,590
PRIOR FILING DATE: 2001-05-04
PRIOR APPLICATION NUMBER: US 60/294,443
PRIOR FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 141
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 126
LENGTH: 829
TYPE: PRT
ORGANISM: Homo sapiens
US-10-058-270A-126

Query Match 100.0%; Score 4369; DB 12; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGLASLLQLQVLCWLOCAASPCRAVFEAEVLEAGGAEOBPGQALGVFGCPG 60
DB 1 MGLPRGLASLLQLQVLCWLOCAASPCRAVFEAEVLEAGGAEOBPGQALGVFGCPG 60

QY 61 QEPALFSTDDNDFTVRNGETVQERRSLKERNPLKIPPSKRILRRHKRDWVAPISVPENG 120
DB 61 QEPALFSTDDNDFTVRNGETVQERRSLKERNPLKIPPSKRILRRHKRDWVAPISVPENG 120

QY 121 KGPFPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLNLKPLDRBEIAK 180
DB 121 KGPFPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLNLKPLDRBEIAK 180

QY 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSLVLEGLPGTSVMQVAT 240
DB 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSLVLEGLPGTSVMQVAT 240

QY 241 DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTIHRSTGTISVSSGLDREKVPYTLTIOA 300
DB 241 DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTIHRSTGTISVSSGLDREKVPYTLTIOA 300

QY 301 TDMGDGSTTTAVAVVEILDANDNAPMFPDQKYEAHVPENAVGEVORLTVTDLAPNSP 360
DB 301 TDMGDGSTTTAVAVVEILDANDNAPMFPDQKYEAHVPENAVGEVORLTVTDLAPNSP 360

QY 361 AWRATYIMGGDDGDHFTTIHPESNQILTRKGLDPEAKNQHTLYVEVNEAPFVLK 420
DB 361 AWRATYIMGGDDGDHFTTIHPESNQILTRKGLDPEAKNQHTLYVEVNEAPFVLK 420

QY 421 PTSTATIIVHVEDVNEAPVFPVPSKVVEQEGIPTEGVCVYTAEDDPKENOKISYRILR 480
DB 421 PTSTATIIVHVEDVNEAPVFPVPSKVVEQEGIPTEGVCVYTAEDDPKENOKISYRILR 480

QY 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMNDGSPPTTGTGLTLTLID 540
DB 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMNDGSPPTTGTGLTLTLID 540

QY 541 VNDHGPVPEPRQITICNQSPVRHLNITDKDLSPTSPFQAQLTDDSDIYWTAEVNEEGD 600
DB 541 VNDHGPVPEPRQITICNQSPVRHLNITDKDLSPTSPFQAQLTDDSDIYWTAEVNEEGD 600

QY 601 TVVLSLKKFLKQDVTYDVHLSLSDHGKNEQLTVIRATVCDCHGHVETCPGPKWGGFILPVL 660
DB 601 TVVLSLKKFLKQDVTYDVHLSLSDHGKNEQLTVIRATVCDCHGHVETCPGPKWGGFILPVL 660

```
QY 661 GAVLALLFLLVLLVLLVKKKIKIPEPLLLPDDTRDNVFFYGGGGGEEEDDYDITQHR 720
Db 661 GAVLALLFLLVLLVLLVKKKIKIPEPLLLPDDTRDNVFFYGGGGGEEEDDYDITQHR 720
QY 721 GLEAPVVLNDVAPTIIPTPMYRPRPANDEIGNFIENLKAANTDPTAPPYDITLLVF 780
Db 721 GLEAPVVLNDVAPTIIPTPMYRPRPANDEIGNFIENLKAANTDPTAPPYDITLLVF 780
QY 781 DYEGSGDAASLSLTSASDQDDYDYLNWGSRFKKLADMYGGGDD 829
Db 781 DYEGSGDAASLSLTSASDQDDYDYLNWGSRFKKLADMYGGGDD 829

RESULT 5
US-10-174-677-21
; Sequence 21, Application US/10174677
; Publication No. US20030190704A1
; GENERAL INFORMATION:
; APPLICANT: Xie, Ting
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ANCHORING STEM CELLS IN A MICROENVIR
; FILE REFERENCE: 40716(IP-012)
; CURRENT APPLICATION NUMBER: US/10/174,677
; PRIOR FILING DATE: 2002-06-19
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-174-677-21

Query Match 100.0%; Score 4369; DB 14; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASLLLLQVLCWQAASEPCRAVFPREAEVTLKAGABQBPQALGVFMGCPG 60
Db 1 MGLPRGPLASLLLLQVLCWQAASEPCRAVFPREAEVTLKAGABQBPQALGVFMGCPG 60
QY 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120
Db 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120
QY 121 KGPPFQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWLNLNKLDRBEEIAK 180
Db 121 KGPPFQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWLNLNKLDRBEEIAK 180
QY 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTPRGSVLEGVLPGTSMQVAT 240
Db 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTPRGSVLEGVLPGTSMQVAT 240
QY 241 DEDDAIYTYNGVWVAYSIIHSQEPKDPHDLMTIHRSTGTISVISGLDREKVPYTLTIOA 300
Db 241 DEDDAIYTYNGVWVAYSIIHSQEPKDPHDLMTIHRSTGTISVISGLDREKVPYTLTIOA 300
QY 301 TDMGDGSGTTTAVAVVAILDANAPMPDPQKYEAHVPENAVGHVORLTVTDLDPNSP 360
Db 301 TDMGDGSGTTTAVAVVAILDANAPMPDPQKYEAHVPENAVGHVORLTVTDLDPNSP 360
QY 361 AWRATYILMGDDGDHDTIITHPESNGQILITRKGLDFAKNQHTLYVEVTNEAPFVKL 420
Db 361 AWRATYILMGDDGDHDTIITHPESNGQILITRKGLDFAKNQHTLYVEVTNEAPFVKL 420
QY 421 PTSTATIVHVEDVNEAPFVPPSKVVEVEGPIPTGEPVCVYTAEDPDKENQKISRIILR 480
Db 421 PTSTATIVHVEDVNEAPFVPPSKVVEVEGPIPTGEPVCVYTAEDPDKENQKISRIILR 480
QY 481 DPAGWLAMPDPSGGQVAVGLDREDEQFVRNNIYVNMVLANDNGSPPTTGTGLLTLLID 540
Db 481 DPAGWLAMPDPSGGQVAVGLDREDEQFVRNNIYVNMVLANDNGSPPTTGTGLLTLLID 540
QY 541 VNDHGVPPEPRQITICNQS FVRHVLNITDKLSPTSFPQALTTDDSDIYWTAEVNEEGD 600
```

```
Db 541 VNDHGVPPEPRQITICNQS FVRHVLNITDKLSPTSFPQALTTDDSDIYWTAEVNEEGD 600
QY 601 TVVLSLKKFLKQDITYDHLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
Db 601 TVVLSLKKFLKQDITYDHLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
QY 661 GAVLALLFLLVLLVLLVKKKIKIPEPLLLPDDTRDNVFFYGGGGGEEEDDYDITQHR 720
Db 661 GAVLALLFLLVLLVLLVKKKIKIPEPLLLPDDTRDNVFFYGGGGGEEEDDYDITQHR 720
QY 721 GLEAPVVLNDVAPTIIPTPMYRPRPANDEIGNFIENLKAANTDPTAPPYDITLLVF 780
Db 721 GLEAPVVLNDVAPTIIPTPMYRPRPANDEIGNFIENLKAANTDPTAPPYDITLLVF 780
QY 781 DYEGSGDAASLSLTSASDQDDYDYLNWGSRFKKLADMYGGGDD 829
Db 781 DYEGSGDAASLSLTSASDQDDYDYLNWGSRFKKLADMYGGGDD 829

RESULT 6
US-10-158-123-1
; Sequence 1, Application US/10158123
; Publication No. US20030194406A1
; GENERAL INFORMATION:
; APPLICANT: Reinhard, Christoph
; APPLICANT: Klingner, Julie
; APPLICANT: Jefferson, Ann
; APPLICANT: Escobedo, Jaime
; APPLICANT: Randazzo, Fillipo
; APPLICANT: Winter, Jill
; APPLICANT: Goodson, Robert
; APPLICANT: Qi, Weimin
; TITLE OF INVENTION: P-Cadherin as a Target for Anti-Cancer
; FILE REFERENCE: 35784/258994
; CURRENT APPLICATION NUMBER: US/10/158,123
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: 60/294,225
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-158-123-1

Query Match 100.0%; Score 4369; DB 14; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASLLLLQVLCWQAASEPCRAVFPREAEVTLKAGABQBPQALGVFMGCPG 60
Db 1 MGLPRGPLASLLLLQVLCWQAASEPCRAVFPREAEVTLKAGABQBPQALGVFMGCPG 60
QY 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120
Db 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKILRRHKRDWVAPISVPENG 120
QY 121 KGPPFQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWLNLNKLDRBEEIAK 180
Db 121 KGPPFQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWLNLNKLDRBEEIAK 180
QY 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTPRGSVLEGVLPGTSMQVAT 240
Db 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTPRGSVLEGVLPGTSMQVAT 240
QY 241 DEDDAIYTYNGVWVAYSIIHSQEPKDPHDLMTIHRSTGTISVISGLDREKVPYTLTIOA 300
Db 241 DEDDAIYTYNGVWVAYSIIHSQEPKDPHDLMTIHRSTGTISVISGLDREKVPYTLTIOA 300
QY 301 TDMGDGSGTTTAVAVVAILDANAPMPDPQKYEAHVPENAVGHVORLTVTDLDPNSP 360
```


Db 301 TDMGDSGTTTAVAVVEILDANDNAPMFDQKYEAHPENAVGHEVQELTVTDLDPNSP 360
Qy 361 AWRATYILINGDDGDHFTTTHPESNOGILTRKGLDPEAKNOHTLYVEVNEAPFVLKL 420
Db 361 AWRATYILINGDDGDHFTTTHPESNOGILTRKGLDPEAKNOHTLYVEVNEAPFVLKL 420
Qy 421 PTSTATIIVHVEDVNEAPVFPSPKVVVEQEGPTGEPVCVYTAEDDPKXENOKISYRILR 480
Db 421 PTSTATIIVHVEDVNEAPVFPSPKVVVEQEGPTGEPVCVYTAEDDPKXENOKISYRILR 480
Qy 481 DPAGWLAMPDPSGQVTVAGTGLDREDEQFVRNNIYEVWVLMNDNGSPPTTGTGLTLLTLD 540
Db 481 DPAGWLAMPDPSGQVTVAGTGLDREDEQFVRNNIYEVWVLMNDNGSPPTTGTGLTLLTLD 540
Qy 541 VNDHGPVPEPRQITTCNOSPVRHVLNITDOLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
Db 541 VNDHGPVPEPRQITTCNOSPVRHVLNITDOLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
Qy 601 TVVLSLKKFLKQDPTDYVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILPVL 660
Db 601 TVVLSLKKFLKQDPTDYVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILPVL 660
Qy 661 GAVLALLFLLVLLVLLVLRKKRKEPILLPEDDTDRDNVYVYGGEGGEDDQDYYITQLHR 720
Db 661 GAVLALLFLLVLLVLLVLRKKRKEPILLPEDDTDRDNVYVYGGEGGEDDQDYYITQLHR 720
Qy 721 GLEARPEVVLNRNDVAPTIIPTPMYRPRANPDEIGNFIIENLKAANTDPTAPPYDILLVF 780
Db 721 GLEARPEVVLNRNDVAPTIIPTPMYRPRANPDEIGNFIIENLKAANTDPTAPPYDILLVF 780
Qy 781 DYEGSGSDAASLSLTSASDQDQDYDLNENWSRFFKKLADMYGGGEDD 829
Db 781 DYEGSGSDAASLSLTSASDQDQDYDLNENWSRFFKKLADMYGGGEDD 829

RESULT 7

US-10-295-027-783
; Sequence 783, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250

; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 783
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-783

Query Match 100.0%; Score 4369; DB 15; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGLPRGSLASILLQVCMLOCAASEPCRAVREAVTLEAGGAEQPCGALCKVFMGCGP 60
Db 1 MGLPRGSLASILLQVCMLOCAASEPCRAVREAVTLEAGGAEQPCGALCKVFMGCGP 60
Qy 61 QBPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDVWAPIVSPENG 120
Db 61 QBPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDVWAPIVSPENG 120
Qy 121 KGPFFQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLILNKLPLDREEIAK 180
Db 121 KGPFFQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLILNKLPLDREEIAK 180
Qy 181 YELFGHVSNGASVEDPMNISIIYTDQNDHKPKTQDTFRGSLVLEGLVPGTSVMQVAT 240
Db 181 YELFGHVSNGASVEDPMNISIIYTDQNDHKPKTQDTFRGSLVLEGLVPGTSVMQVAT 240
Qy 241 DEDDAIYTVNGVAVYSIHSQEPKDPHLMFTIHRSTGTISVISSGLDREKVEPYTLTQA 300
Db 241 DEDDAIYTVNGVAVYSIHSQEPKDPHLMFTIHRSTGTISVISSGLDREKVEPYTLTQA 300
Qy 301 TMDGDSGTTTAVAVVEILDANDNAPMFDQKYEAHPENAVGHEVQELTVTDLDPNSP 360
Db 301 TMDGDSGTTTAVAVVEILDANDNAPMFDQKYEAHPENAVGHEVQELTVTDLDPNSP 360
Qy 361 AWRATYILINGDDGDHFTTTHPESNOGILTRKGLDPEAKNOHTLYVEVNEAPFVLKL 420
Db 361 AWRATYILINGDDGDHFTTTHPESNOGILTRKGLDPEAKNOHTLYVEVNEAPFVLKL 420
Qy 421 PTSTATIIVHVEDVNEAPVFPSPKVVVEQEGPTGEPVCVYTAEDDPKXENOKISYRILR 480
Db 421 PTSTATIIVHVEDVNEAPVFPSPKVVVEQEGPTGEPVCVYTAEDDPKXENOKISYRILR 480
Qy 481 DPAGWLAMPDPSGQVTVAGTGLDREDEQFVRNNIYEVWVLMNDNGSPPTTGTGLTLLTLD 540
Db 481 DPAGWLAMPDPSGQVTVAGTGLDREDEQFVRNNIYEVWVLMNDNGSPPTTGTGLTLLTLD 540
Qy 541 VNDHGPVPEPRQITTCNOSPVRHVLNITDOLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
Db 541 VNDHGPVPEPRQITTCNOSPVRHVLNITDOLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
Qy 601 TVVLSLKKFLKQDPTDYVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILPVL 660
Db 601 TVVLSLKKFLKQDPTDYVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFILPVL 660
Qy 661 GAVLALLFLLVLLVLLVLRKKRKEPILLPEDDTDRDNVYVYGGEGGEDDQDYYITQLHR 720
Db 661 GAVLALLFLLVLLVLLVLRKKRKEPILLPEDDTDRDNVYVYGGEGGEDDQDYYITQLHR 720
Qy 721 GLEARPEVVLNRNDVAPTIIPTPMYRPRANPDEIGNFIIENLKAANTDPTAPPYDILLVF 780
Db 721 GLEARPEVVLNRNDVAPTIIPTPMYRPRANPDEIGNFIIENLKAANTDPTAPPYDILLVF 780
Qy 781 DYEGSGSDAASLSLTSASDQDQDYDLNENWSRFFKKLADMYGGGEDD 829
Db 781 DYEGSGSDAASLSLTSASDQDQDYDLNENWSRFFKKLADMYGGGEDD 829

RESULT 8
US-10-295-027-800
; Sequence 800, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-01250005
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 800
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-800

Query Match 100.0%; Score 4369; DB 15; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASLLQLQVCLWCAASEPCRAVFAEVLTEAGGAPQEPGQALGVFMGCPG 60
DB 1 MGLPRGPLASLLQLQVCLWCAASEPCRAVFAEVLTEAGGAPQEPGQALGVFMGCPG 60

QY 61 QEPALFSTNDNDFVRNGETVOERSLKERPLKIFSKRLRHRKEDWVAPISVPENG 120
DB 61 QEPALFSTNDNDFVRNGETVOERSLKERPLKIFSKRLRHRKEDWVAPISVPENG 120

QY 121 KGPPFQRLNQLKSKNRDRTKIFYSTIGPADSPPEGVFAVEKETGWLKLLNKLDRREEIAK 180
DB 121 KGPPFQRLNQLKSKNRDRTKIFYSTIGPADSPPEGVFAVEKETGWLKLLNKLDRREEIAK 180

QY 181 YELFGHVAENGASVEDPMNIIIVTDNDHKPKFTQDTRGSLVLEGVLPGTSVMQVAT 240
DB 181 YELFGHVAENGASVEDPMNIIIVTDNDHKPKFTQDTRGSLVLEGVLPGTSVMQVAT 240

QY 241 DEDDAIYTYNGVWYSHSQBPDPKPHDLMFTIHRSTGTISVISSGLDREKVPYTLTIOA 300
DB 241 DEDDAIYTYNGVWYSHSQBPDPKPHDLMFTIHRSTGTISVISSGLDREKVPYTLTIOA 300

QY 301 TDMGDGGSTTTAVAVVEILDANDNAPMEDPQKYEAHVPENAVAGHEVQRLTVDLDAPNSP 360
DB 301 TDMGDGGSTTTAVAVVEILDANDNAPMEDPQKYEAHVPENAVAGHEVQRLTVDLDAPNSP 360

QY 361 AWRATYLMGDDGDGHFTITTHPESNQILTRKGLDPEAKNQHTLYVEVNEAPFVLKL 420
DB 361 AWRATYLMGDDGDGHFTITTHPESNQILTRKGLDPEAKNQHTLYVEVNEAPFVLKL 420

QY 421 PTSTATIVVHVEDVNEAPVFPVPPSKVVEQEGIPTEBPVCVVTAEDDPKQKISYRILR 480
DB 421 PTSTATIVVHVEDVNEAPVFPVPPSKVVEQEGIPTEBPVCVVTAEDDPKQKISYRILR 480

QY 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMDNQSPPTGTGTLTLLID 540
DB 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMDNQSPPTGTGTLTLLID 540

QY 541 VNDHGPVPEPROITICNOSPVRHVNLITDKLSHPTSPFOQLTDDSDIYWTAEVNEEGD 600
DB 541 VNDHGPVPEPROITICNOSPVRHVNLITDKLSHPTSPFOQLTDDSDIYWTAEVNEEGD 600

QY 601 TVVLSLKKFLKQDITYDVHLSLSHGKNEQLTVIRATVCDCHGVETCFGPWKGFILPVL 660
DB 601 TVVLSLKKFLKQDITYDVHLSLSHGKNEQLTVIRATVCDCHGVETCFGPWKGFILPVL 660

QY 661 GAVTALLFLLVLLVLRKKIKIPELLPEDDTRDNVFFYGEBSGSEEDODYDITOLHR 720
DB 661 GAVTALLFLLVLLVLRKKIKIPELLPEDDTRDNVFFYGEBSGSEEDODYDITOLHR 720

QY 721 GLEARPEVVLNRDVAPTIPTPMVPRPANDEIGNFIENLKAANTDFTAPPYDTLLVF 780
DB 721 GLEARPEVVLNRDVAPTIPTPMVPRPANDEIGNFIENLKAANTDFTAPPYDTLLVF 780

QY 781 DYEAGSGDAASLSLTSASDQDQDYVLYNEWGRFVKLADMYGGEDD 829
DB 781 DYEAGSGDAASLSLTSASDQDQDYVLYNEWGRFVKLADMYGGEDD 829

RESULT 9
US-10-295-027-863
; Sequence 863, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-01250005
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10

;; PRIOR APPLICATION NUMBER: US 60/355,250
;; PRIOR FILING DATE: 2002-02-08
;; PRIOR APPLICATION NUMBER: US 60/356,714
;; PRIOR FILING DATE: 2002-02-13
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 1386
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 863
;; LENGTH: 829
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-295-027-863

Query Match 100.0%; Score 4369; DB 15; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGLPRGPLASLLLLQVLCWLCQAASEPCRAVFREAEVTLAAGGAEQPGQALGVFMGCPG 60
Db 1 MGLPRGPLASLLLLQVLCWLCQAASEPCRAVFREAEVTLAAGGAEQPGQALGVFMGCPG 60

Qy 61 QEPALFSTDDNDFTVRNGETVOERSLKERNPLKIFPSKILRRHKEDWVAPISVPENG 120
Db 61 QEPALFSTDDNDFTVRNGETVOERSLKERNPLKIFPSKILRRHKEDWVAPISVPENG 120

Qy 121 KGPPQRLNQLKSKNRDRTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREEIAK 180
Db 121 KGPPQRLNQLKSKNRDRTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREEIAK 180

Qy 181 YELFGHVAIVNGVAYSIIHSGQKDEKDHLMFTIHRSTGTISVISGLDREKVPYTLTQA 300
Db 181 YELFGHVAIVNGVAYSIIHSGQKDEKDHLMFTIHRSTGTISVISGLDREKVPYTLTQA 300

Qy 241 DEDDAIYTYNGVAYSIIHSGQKDEKDHLMFTIHRSTGTISVISGLDREKVPYTLTQA 300
Db 241 DEDDAIYTYNGVAYSIIHSGQKDEKDHLMFTIHRSTGTISVISGLDREKVPYTLTQA 300

RESULT 10
US-10-295-027-896
;; Sequence 896, Application US/10295027
;; Publication NO. US20030232350A1
;; GENERAL INFORMATION:
;; APPLICANT: Afar, Daniel
;; APPLICANT: Aziz, Natasha
;; APPLICANT: Ginsberg, Wendy M.
;; APPLICANT: Gish, Kurt C.
;; APPLICANT: Glynn, Richard
;; APPLICANT: Hevezi, Peter A.
;; APPLICANT: Mack, David H.
;; APPLICANT: Murray, Richard
;; APPLICANT: Watson, Susan R.
;; APPLICANT: Eos Biotechnology, Inc.
;; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
;; FILE REFERENCE: 018501-01250005
;; CURRENT APPLICATION NUMBER: US/10/295,027
;; CURRENT FILING DATE: 2002-11-13
;; PRIOR APPLICATION NUMBER: US 09/663,733
;; PRIOR FILING DATE: 2000-09-15
;; PRIOR APPLICATION NUMBER: US 60/350,666
;; PRIOR FILING DATE: 2001-11-13
;; PRIOR APPLICATION NUMBER: US 60/335,394
;; PRIOR FILING DATE: 2001-11-15
;; PRIOR APPLICATION NUMBER: US 60/332,464
;; PRIOR FILING DATE: 2001-11-21
;; PRIOR APPLICATION NUMBER: US 60/334,393
;; PRIOR FILING DATE: 2001-11-29
;; PRIOR APPLICATION NUMBER: US 60/340,376
;; PRIOR FILING DATE: 2001-12-14
;; PRIOR APPLICATION NUMBER: US 60/347,211
;; PRIOR FILING DATE: 2002-01-08
;; PRIOR APPLICATION NUMBER: US 60/347,349
;; PRIOR FILING DATE: 2002-01-10
;; PRIOR APPLICATION NUMBER: US 60/355,250
;; PRIOR FILING DATE: 2002-02-08
;; PRIOR APPLICATION NUMBER: US 60/356,714
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 1386
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 896
;; LENGTH: 829
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-295-027-896

Query Match 100.0%; Score 4369; DB 15; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGLPRGPLASLLLLQVLCWLCQAASEPCRAVFREAEVTLAAGGAEQPGQALGVFMGCPG 60
Db 1 MGLPRGPLASLLLLQVLCWLCQAASEPCRAVFREAEVTLAAGGAEQPGQALGVFMGCPG 60

Qy 61 QEPALFSTDDNDFTVRNGETVOERSLKERNPLKIFPSKILRRHKEDWVAPISVPENG 120
Db 61 QEPALFSTDDNDFTVRNGETVOERSLKERNPLKIFPSKILRRHKEDWVAPISVPENG 120

Qy 121 KGPPQRLNQLKSKNRDRTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREEIAK 180
Db 121 KGPPQRLNQLKSKNRDRTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREEIAK 180

Qy 181 YELFGHVAIVNGVAYSIIHSGQKDEKDHLMFTIHRSTGTISVISGLDREKVPYTLTQA 240
Db 181 YELFGHVAIVNGVAYSIIHSGQKDEKDHLMFTIHRSTGTISVISGLDREKVPYTLTQA 240

Qy 241 DEDDAIYTYNGVAYSIIHSGQKDEKDHLMFTIHRSTGTISVISGLDREKVPYTLTQA 300
Db 241 DEDDAIYTYNGVAYSIIHSGQKDEKDHLMFTIHRSTGTISVISGLDREKVPYTLTQA 300

RESULT 14
 US-10-295-027-338
 ; Sequence 338, Application US/10295027
 ; Publication No. US20030232350A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Afar, Daniel
 ; APPLICANT: Ginsberg, Wendy M.
 ; APPLICANT: Gish, Kurt C.
 ; APPLICANT: Glynne, Richard
 ; APPLICANT: Hevezi, Peter A.
 ; APPLICANT: Mack, David H.
 ; APPLICANT: Murray, Richard
 ; APPLICANT: Watson, Susan R.
 ; APPLICANT: Eos Biotechnology, Inc.
 ; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
 ; FILE REFERENCE: 018501-012500US
 ; CURRENT APPLICATION NUMBER: US/10/295,027
 ; CURRENT FILING DATE: 2002-11-13
 ; PRIOR APPLICATION NUMBER: US 09/663,733
 ; PRIOR FILING DATE: 2000-09-15
 ; PRIOR APPLICATION NUMBER: US 60/350,666
 ; PRIOR FILING DATE: 2001-11-13
 ; PRIOR APPLICATION NUMBER: US 60/335,394
 ; PRIOR FILING DATE: 2001-11-15
 ; PRIOR APPLICATION NUMBER: US 60/332,464
 ; PRIOR FILING DATE: 2001-11-21
 ; PRIOR APPLICATION NUMBER: US 60/334,393
 ; PRIOR FILING DATE: 2001-11-29
 ; PRIOR APPLICATION NUMBER: US 60/340,376
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: US 60/347,211
 ; PRIOR FILING DATE: 2002-01-08
 ; PRIOR APPLICATION NUMBER: US 60/347,349
 ; PRIOR FILING DATE: 2002-01-10
 ; PRIOR APPLICATION NUMBER: US 60/355,250
 ; PRIOR FILING DATE: 2002-02-08
 ; PRIOR APPLICATION NUMBER: US 60/356,714
 ; PRIOR FILING DATE: 2002-02-13
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 1386
 ; SOFTWARE: PatentIn ver. 2.1
 ; SEQ ID NO 338
 ; LENGTH: 829
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-295-027-338

Query Match 99.8%; Score 4361; DB 15; Length 829;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MGLPRGLASLLILQVLCWQCAASEPCRAVREAEVTLKAGGAEQFGQALGKVFVFCG 60
Db	1	MGLPRGLASLLILQVLCWQCAASEPCRAVREAEVTLKAGGAEQFGQALGKVFVFCG 60
Qy	61	QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKILRRHRKRDWVAPISVPENG 120
Db	61	QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKILRRHRKRDWVAPISVPENG 120
Qy	121	KGPPQSLNOLKSNKDRDKIFYSITCPGADSPPEGVFAVEKETGMLLNKPLDREIAK 180
Db	121	KGPPQSLNOLKSNKDRDKIFYSITCPGADSPPEGVFAVEKETGMLLNKPLDREIAK 180
Qy	181	YELFGHAVSENGASVEDPMNISITVDQNDHKPKFTQDTRFGSVLGGVPGTSMQVAT 240
Db	181	YELFGHAVSENGASVEDPMNISITVDQNDHKPKFTQDTRFGSVLGGVPGTSMQVAT 240
Qy	241	DEDAIYTVNGVAYSVTHSQEPKDPHDLMTIHRSTGTISVSSGLDREKVPYTLTICA 300
Db	241	DEDAIYTVNGVAYSVTHSQEPKDPHDLMTIHRSTGTISVSSGLDREKVPYTLTICA 300

Qy	301	TMDDGSGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHVORLTVTDLDAENSP 360
Db	301	TMDDGSGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHVORLTVTDLDAENSP 360
Qy	361	AWRATYILMGDDGDGHFTITTHPESNQILTRKGLDPEAKNQHTLYVEVTNEAPFVYKL 420
Db	361	AWRATYILMGDDGDGHFTITTHPESNQILTRKGLDPEAKNQHTLYVEVTNEAPFVYKL 420
Qy	421	PTSTATIVVHVEDVNEAPVFPVPSKVVEVQEGIPTEGFCVYTAEDDCKENKISYILR 480
Db	421	PTSTATIVVHVEDVNEAPVFPVPSKVVEVQEGIPTEGFCVYTAEDDCKENKISYILR 480
Qy	481	DPAGWLAMPDPSGOVTAAGTLDRDEQFVRNNIYEVVWVLANMDSPPPTGTGTLILLID 540
Db	481	DPAGWLAMPDPSGOVTAAGTLDRDEQFVRNNIYEVVWVLANMDSPPPTGTGTLILLID 540
Qy	541	VNDHGPVPEPQITICNQSPVRVHLNITDKLSPTSPFQALQTLTDDSDIYWTABVNEEG 600
Db	541	VNDHGPVPEPQITICNQSPVRVHLNITDKLSPTSPFQALQTLTDDSDIYWTABVNEEG 600
Qy	601	TVVLSLKKFLKQDTYDVHLSLSHDHNGEQLTVIRATVCDCHGHVETCPGPKWGGFILPVL 660
Db	601	TVVLSLKKFLKQDTYDVHLSLSHDHNGEQLTVIRATVCDCHGHVETCPGPKWGGFILPVL 660
Qy	661	GAVLALLFLLVLLVLRKKIKIPELPLLPEDDTRDNVYFYEGEGGEDDQDYDITQLHR 720
Db	661	GAVLALLFLLVLLVLRKKIKIPELPLLPEDDTRDNVYFYEGEGGEDDQDYDITQLHR 720
Qy	721	GLEARPEVVLRENDVAPTIIPTPMYRPRPANPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
Db	721	GLEARPEVVLRENDVAPTIIPTPMYRPRPANPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
Qy	781	DYEGSGSDAASLSLTSASDQDQDYDLNEMWGSFRFKKLADMYGGGDD 829
Db	781	DYEGSGSDAASLSLTSASDQDQDYDLNEMWGSFRFKKLADMYGGGDD 829

RESULT 15
 US-10-295-027-1246
 ; Sequence 1246, Application US/10295027
 ; Publication No. US20030232350A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Afar, Daniel
 ; APPLICANT: Aziz, Natasha
 ; APPLICANT: Ginsberg, Wendy M.
 ; APPLICANT: Gish, Kurt C.
 ; APPLICANT: Glynne, Richard
 ; APPLICANT: Hevezi, Peter A.
 ; APPLICANT: Mack, David H.
 ; APPLICANT: Murray, Richard
 ; APPLICANT: Watson, Susan R.
 ; APPLICANT: Eos Biotechnology, Inc.
 ; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
 ; FILE REFERENCE: 018501-012500US
 ; CURRENT APPLICATION NUMBER: US/10/295,027
 ; CURRENT FILING DATE: 2002-11-13
 ; PRIOR APPLICATION NUMBER: US 09/663,733
 ; PRIOR FILING DATE: 2000-09-15
 ; PRIOR APPLICATION NUMBER: US 60/350,666
 ; PRIOR FILING DATE: 2001-11-13
 ; PRIOR APPLICATION NUMBER: US 60/335,394
 ; PRIOR FILING DATE: 2001-11-15
 ; PRIOR APPLICATION NUMBER: US 60/332,464
 ; PRIOR FILING DATE: 2001-11-21
 ; PRIOR APPLICATION NUMBER: US 60/334,393
 ; PRIOR FILING DATE: 2001-11-29
 ; PRIOR APPLICATION NUMBER: US 60/340,376
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: US 60/347,211
 ; PRIOR FILING DATE: 2002-01-08
 ; PRIOR APPLICATION NUMBER: US 60/347,349
 ; PRIOR FILING DATE: 2002-01-10

Search completed: September 21, 2004, 22:20:46
Job time : 147 secs

```
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1246
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-1246

Query Match      99.8%; Score 4361; DB 15; Length 829;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 MGLPRGLASLLLOVCMQLQCAASEPCRAVFEAEVTLAAGAEQEPGQALGVFMGCGP 60
Db      1 MGLPRGLASLLLOVCMQLQCAASEPCRAVFEAEVTLAAGAEQEPGQALGVFMGCGP 60
QY      61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVVAPIISVPENG 120
Db      61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVVAPIISVPENG 120
QY      121 KGFPFQRLNQLKSNKDRDKIFYSITGPGADSPGEGFAVEKETGWLILNKLPLDREETAK 180
Db      121 KGFPFQRLNQLKSNKDRDKIFYSITGPGADSPGEGFAVEKETGWLILNKLPLDREETAK 180
QY      181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDFRGSVLGVLPGTSVMQVAT 240
Db      181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDFRGSVLGVLPGTSVMQVAT 240
QY      241 DEDDAIYTVNGVAVSIHSGEPKDPHDLMTIHRSTGTISVISSSGLDREKVPYTLTIOA 300
Db      241 DEDDAIYTVNGVAVSIHSGEPKDPHDLMTIHRSTGTISVISSSGLDREKVPYTLTIOA 300
QY      301 TMDMGGSSTTAVAVVEILDANDNAPMFDPKQYEAHVPENAVGEVQRLTVDLDAPNSP 360
Db      301 TMDMGGSSTTAVAVVEILDANDNAPMFDPKQYEAHVPENAVGEVQRLTVDLDAPNSP 360
QY      361 AWRATVLMGDDGDGHTIITHPESNQGLITTRKGLDFAKNQHTLYVEVTNEAPFVLKL 420
Db      361 AWRATVLMGDDGDGHTIITHPESNQGLITTRKGLDFAKNQHTLYVEVTNEAPFVLKL 420
QY      421 PTSTATIIVHVEDVNEAPVFPSPKVVVEQEGIPTGECVCTAEDPDKENQKISYRIILR 480
Db      421 PTSTATIIVHVEDVNEAPVFPSPKVVVEQEGIPTGECVCTAEDPDKENQKISYRIILR 480
QY      481 DPAGWLAMDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGTLTLTLD 540
Db      481 DPAGWLAMDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGTLTLTLD 540
QY      541 VNDHGPVPEPRQITICNQSPVRHVLNITDKDLSFHTSPFOAQLTDDSDIYWTAEVNEEGD 600
Db      541 VNDHGPVPEPRQITICNQSPVRHVLNITDKDLSFHTSPFOAQLTDDSDIYWTAEVNEEGD 600
QY      601 TVVLSLKKFLKQDYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL 660
Db      601 TVVLSLKKFLKQDYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL 660
QY      661 GAVLALLFLVLLVLLVLRKKRIKEPLLALPEDDTRDNVFIYEGEGGEEQDYITQLHR 720
Db      661 GAVLALLFLVLLVLLVLRKKRIKEPLLALPEDDTRDNVFIYEGEGGEEQDYITQLHR 720
QY      721 GLEARPEVLNDVAPTIIPTMYRPRPANDPDEIGNFIIENLKAANTDPTAPPYDILLVF 780
Db      721 GLEARPEVLNDVAPTIIPTMYRPRPANDPDEIGNFIIENLKAANTDPTAPPYDILLVF 780
QY      781 DYEGSGSDAASLSLTSASDQDDYDYLNEGWSRFFKLLADMYGGGEDD 829
Db      781 DYEGSGSDAASLSLTSASDQDDYDYLNEGWSRFFKLLADMYGGGEDD 829
```

This Page Blank (usp10)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 23, 2004, 03:15:06 ; Search time 5081 Seconds
(without alignments)
4872.220 Million cell updates/sec

Title: US-09-916-849A-1
Perfect score: 829
Sequence: 1 MGLPRGPLASLLLLQVCWLQ.....NEWGSRFKKLADMYGGGDD 829

Scoring table:
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Word size: 1

Total number of hits satisfying chosen parameters: 55025477

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=framed p2n.model -DEV=xlh
-O=/cpn2.1/uspto.spool.h/us09916849/runat_21092004.145835.7328/app.query.fasta_1.967
-DB=EST -QFMT=fastap -SUFFIX=oligo.rst -MINMATCH=0 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09916849 @CGN 1.1 3609 @runat_21092004.145835.7328 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_nam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_prg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	659	79.5	3632	11	BC014462	BC014462 Homo sapi
2	243	29.3	913	13	BQ676796	BQ676796 AGENCOURT
3	238	28.7	918	14	CA454560	CA454560 AGENCOURT
4	235	28.3	1016	12	BM923882	BM923882 AGENCOURT
5	228	27.5	875	14	CA489015	CA489015 AGENCOURT
6	228	27.5	1122	12	BM545820	BM545820 AGENCOURT
7	227	27.4	769	12	BG403092	BG403092 602418911
8	224	27.0	887	14	CA455046	CA455046 AGENCOURT
9	221	26.7	865	13	BUS43746	BUS43746 AGENCOURT
10	214	25.8	703	12	BM715802	BM715802 UI-E-EJO-
11	214	25.8	1006	13	BUI68948	BUI68948 AGENCOURT
12	213	25.7	745	14	CB240552	CB240552 UI-CF-FNO
13	211	25.5	880	14	CB922509	CB922509 AGENCOURT
14	211	25.5	894	14	CA489124	CA489124 AGENCOURT
15	208	25.1	999	12	BM804940	BM804940 AGENCOURT
16	204	24.6	836	12	BM045728	BM045728 603624563
17	200	24.1	846	13	BQ690885	BQ690885 AGENCOURT
18	197	23.8	892	14	CA489338	CA489338 AGENCOURT
19	196	23.6	815	14	CB997053	CB997053 AGENCOURT
20	196	23.6	835	14	CB996355	CB996355 AGENCOURT
21	196	23.6	949	13	BQ679513	BQ679513 AGENCOURT
22	195	23.5	899	14	CA454360	CA454360 AGENCOURT
23	180	21.7	854	14	CA488032	CA488032 AGENCOURT
24	167	20.1	1040	12	BM909817	BM909817 AGENCOURT
25	166	20.0	778	12	BM049731	BM049731 603624246
26	160	19.3	1001	13	BQ072354	BQ072354 AGENCOURT
27	151	18.2	475	10	BF915077	BF915077 IL3-UT011
28	151	18.1	873	14	CA453911	CA453911 AGENCOURT
29	150	18.1	773	10	BE740340	BE740340 601594641
30	149	18.0	674	9	AU141604	AU141604 601594641
31	149	18.0	902	14	CB994831	CB994831 AGENCOURT
32	144	17.4	433	12	BQ62473	BQ62473 IL3-UT011
33	142	17.1	912	13	BQ928283	BQ928283 AGENCOURT
34	141	17.0	611	10	AW950190	AW950190 EST382155
35	140	16.9	927	14	CA489357	CA489357 AGENCOURT
36	135	16.3	901	14	CA454213	CA454213 AGENCOURT
37	133	16.0	870	10	BE378969	BE378969 601237596
38	131	15.8	916	13	BUI79183	BUI79183 AGENCOURT
39	126	15.2	848	12	BG675452	BG675452 602621781
40	122	14.7	888	14	CA487856	CA487856 AGENCOURT
41	122	14.7	1049	13	BQ936686	BQ936686 AGENCOURT
42	121	14.6	364	12	BG319601	BG319601 4dh29 Hum
43	118	14.2	761	10	BE741190	BE741190 601593814
44	117	14.1	790	14	CF593464	CF593464 AGENCOURT
45	117	14.1	878	14	CA489425	CA489425 AGENCOURT

ALIGNMENTS

RESULT 1
BC014462
LOCUS BC014462
DEFINITION Homo sapiens cadherin 3, type 1, P-cadherin (placental), mRNA (cdna clone IMAGE:4870356), containing frame-shift errors.
ACCESSION BC014462
VERSION BC014462.1 GI:15680220
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3632)

Db 1408 GCAGTGGGCGCATGAGTGCAGAGGCTGACGGTCACTGATCTGGACGCCGCCCACTCACCA 1467
Qy 361 AlaTpaAtaAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db 1468 GGTGGCGTGCACCTTACCTTATCATATGGCGGTGACGACGGGACCATTTTACCAATCAC 1527
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1528 ACCACCCCTGAGAGCAACACAGGCGCTCTCTCAACACAGGAGGGTTTGGATTTTGGAGCC 1587
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1588 AAAAACCCAGCACACCTGTACGTGAAGTGACCAAGAGGCCCTTTTGTGTGAAGCTC 1647
Qy 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
Db 1648 CCAACCTCCACAGCCACCATAGTGTCCACGTGGAGGATGTGAATGAGGCACCTGTGTT 1707
Qy 441 ValProProSerLysValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1708 GTCCCAACCCCTCCAAAGTCTGAGTCCAGGAGGCGATCCCACTGGGAGGCTGTGT 1767
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1768 GTCTACACTGCAGAGACCCCTGCACAGGAGATCAAGATCAGCTACCGCATCTCGAGA 1827
Qy 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1828 GACCCAGCAGGGTGGTGTAGCTAGCCATGACCCAGACAGTGGCAGGTCAACAGCTGGGCCACC 1887
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1889 CTCGACCGTGGAGATGAGCAGTTGTGGAGAACCAACATCTATGAAGTATGGTCTTGCC 1947
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuThrLeuIleAsp 540
Db 1948 ATGGCAATGTAAGCCCTCCCACTGCGACCGGAACCCCTCTGCTAACTGATTGAT 2007
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 2008 GTCAACAGCATGGCCCATGCTGAGCCCTGATGATCAGATCAGCATCTGCAACCAAGCCCT 2067
Qy 561 ValArgHis-ValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 2068 GTGGCGCA-GTGTCTGAACATACGGACACAGACCTGTCTCCCACTCCCTCCCTTTCCA 2126
Qy 580 nlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAs 600
Db 2127 GGCCCACTCAGATGACTCAGACATCTACTGCGAGGAGGTCAACAGGAAGGTGA 2186
Qy 600 pThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSe 620
Db 2187 CACAGTGGTCTTGTCTCCCTGAAGAAGTCTCTGAAGCAGGATACATATGAGTGCACCTTTC 2246
Qy 620 rLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCy 640
Db 2247 TCTGTCTGACCATGGCAACAAAGAGCAGCTGACGGTGTATCAGGGCCACTGTGTGGAGCTG 2306
Qy 640 shisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValle 660
Db 2307 CCATGGCCATGTGAAACCTGCCCTGGACCCCTGGAAAGGAGGTTCATCTCCCTCTGTCT 2366
Qy 660 uGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2367 GGGGGCTGTCTGTCTGTCTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2426
Qy 680 sArgLysIleLysGluProLeuLeuLeuProGluLeuAspThrArgAspAsnValPheTy 700
Db 2427 GCGGAAGATCAAGAGCCCTCTCTACTCCAGAAGATGACACCCGCTGACACAGCTCTTCTA 2486
Qy 700 rTyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisAr 720
Db 2487 CTATGGCGAGAGGGGGTGGCGAGAGGACCGAGCTATGACATCACCACCTCCACCG 2546

Qy 720 gGlyLeuGluAlaArgProGluValValLeuAsnAspValAlaProThrIleIlePr 740
Db 2547 AGTCTGAGGGCGAGCGGAGGTGTTCTCCGAATGACGTGGCCACCAACCATCATCCC 2606
Qy 740 oThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPhelleIleG1 760
Db 2607 GACACCCATGTACCGTCTAGCGCAGCAACCCAGATGAATCGCAACTTTTATAATGA 2666
Qy 760 u 760
Db 2667 G 2667
RESULT 2
BQ676796 913 bp mRNA linear EST 15-JUL-2002
LOCUS AGENCOURT 8184299 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6264044
DEFINITION 5' mRNA Sequence.
ACCESSION BQ676796
VERSION BQ676796.1 GI:21789475
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 913)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: Agencourt Bioscience Corporation
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM2428 row: o column: 21
High quality sequence stop: 635.
Location/Qualifiers
1. 913
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6264044"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_112"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.: 9,49e-208 Length: 913
Score: 243.00 Matches: 243
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 29.31% Indels: 0
Gaps: 0
US-09-916-849A-1 (1-829) x BQ676796 (1-913)

Qy 275 SerThrGlyThrIleSerValIleSerSerGlyLeuAspArgGlyLysValProGluTyr 294
Db 2 AGCACAGGCACCATCAGCGTCATCTCCAGTGGCTGGACCGGGAAAAAGTCCCTGAGTAC 61

Qy 295 ThrLeuThrIleGlnAlaThrAspMetAspGlySerThrThrThrAlaValAla 314
 Db 62 ACACCTGACCATCCAGGCCACAGACATGGATGGGACCGCTCCACACACCGCAGTGGCA 121
 Qy 315 ValValGluIleLeuAspAlaAsnAlaProMetPheAspProGlnLysThrGlu 334
 Db 122 GTAGTGGAGATCTTGTATGCCAATGACATGCTCCCATGTTTGATCCCGGAGATACGAG 181
 Qy 335 AlaHisValProGluAsnAlaValGlyHisGluValGlnAlaThrValThrAspLeu 354
 Db 182 GCCCATGTGCTGAGAATGAGTGGGCCCATGAGTGCAGAGGCTGACGGTCACTGATCTG 241
 Qy 355 AspAlaProAsnSerProAlaThrArgAlaThrThrLeuIleMetGlyAspGly 374
 Db 242 GACGCCCCCAACTCACAGCGTGGCGTGCACCTACTTATCATGGCGGTGACGACGGG 301
 Qy 375 AspHisPheThrIleThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLys 394
 Db 302 GACCAITTTACCATACCCACCCCTGAGAGCAACAGGGCATCTGACACACAGGAG 361
 Qy 395 GlyLeuAspPheGluAlaLysAsnGlnHisThrLeuThrValGluValThrAsnGluAla 414
 Db 362 GGTTTGGATTTTGGAGGCCAAACACACGACACCCCTGTACGTTTGAAGTGACCAACAGGCC 421
 Qy 415 PropheValLeuLysLeuProThrSerThrAlaThrIleValValHisValGluAspVal 434
 Db 422 CTTTGTGCTGAGATCTCCACCTCACAGCCACCAATAGTGGTCCACGTGGAGATGTG 481
 Qy 435 AsnGluAlaProValPheValProSerLysValValGluValGlnGlyIlePro 454
 Db 482 AATGAGGCACCTGTGTTTGTCCACCCCTCCAAAGTCGTTGAGGTCCAGGAGGCATCCCC 541
 Qy 455 ThrGlyGluProValCysValThrAlaGluAspProAspLysGluAsnGlnLysIle 474
 Db 542 ACTGGGGAGCCTGTGTGTCTACATCGCAGAGAGCCCTGCAAGAGAGATCAAAAGATC 601
 Qy 475 SerTyrArgIleLeuArgAspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGln 494
 Db 602 AGCTACCGCATCTGAGAGACCCAGCAGGCTGGTAGCATGGACCCAGACAGTGGCAG 661
 Qy 495 ValThrAlaValGlyThrLeuAspArgGluAspGluGlnPheValArgAsnLysIleTyr 514
 Db 662 GTCACAGCTGTGGGACCCCTCGACCGTGGAGTGGAGCAGTGTGTGGAGGAACAATCTAT 721
 Qy 515 GluValMet 517
 Db 722 GAAGTCAATG 730

RESULT 3
 CA454560
 LOCUS
 DEFINITION AGENCOURT 10714108 MAPcL Homo sapiens cDNA clone IMAGE:6720636 5',
 mRNA sequence.
 CA454560
 CA454560.1 GI:24904418
 EST.
 Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 918)
 NIH-MGC <http://mgc.ncl.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Kristi A. Eglund, Ira Pastan
 CDNA Library Prepared by: Invitrogen Corp
 CDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: L1AM14280 row: h column: 12
 High quality sequence stop: 650.
 Location/Qualifiers
 1. 918
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6720636"
 /cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
 hTERT-HME1, LNCaP"
 /lab_host="EMDH10B"
 /clone_lib="MAPcL"
 /notes="Vector: pCMV-SPORT6; Site_1: EcoRV; Site_2: Not I;
 Subtracted with brain, liver, lung, kidney and muscle.
 Directionally cloned. Priming method: oligo-dT. Average
 insert size: 1800 bp. Library amplification: 26,000 fold.
 Kristi A. Eglund, James J. Vincent, Robert Strausberg,
 Bungkok Lee & Ira Pastan: Discovery of new breast
 cancer genes encoding membrane and secreted proteins.
 Manuscript submitted."

ORIGIN
 Alignment Scores:
 Pred. No.: 3,01e-203 Length: 918
 Score: 238.00 Matches: 251
 Percent Similarity: 99.60% Conservative: 0
 Best Local Similarity: 99.60% Mismatches: 0
 Query Match: 28.71% Indels: 1
 DB: 14 Gaps: 0
 US-09-916-849A-1 (1-829) x CA454560 (1-918)

Qy 180 LysTyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMet 199
 Db 3 AAGTATGAGTCTTTGGCCACGCTGTGTACAGAAATGGTGGCTCAGTGAGGAGCCCATG 62
 Qy 200 AsnIleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThr 219
 Db 63 AACATCTCCATCATCGTGACCGACCAAGATGACCAAGCCCAAGTTTACCCAGGACAC 122
 Qy 220 PheArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAla 239
 Db 123 TTCGAGGGAGTGTCTTAGAGGGAGTCTTACCAGGACTTCTGTGATGAGGAGACGCC 182
 Qy 240 ThrAspGluAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSer 259
 Db 183 ACGGATGAGGATGATGCACTACCTACCAATGGGGTGGTTCCTTACTCCATCCATAGC 242
 Qy 260 GlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIle 279
 Db 243 CAAGAACAAGACACCCACACGACCTCATGTCCACATTACCCGAGACAGCACCATC 302
 Qy 280 SerValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGln 299
 Db 303 AGCGTCACTCTCCAGTGGCTGGACCGGAAAGTCCCTGAGTACACACTGACCATCCAG 362
 Qy 300 AlaThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluIleLeu 319
 Db 363 GCCACAGACATGGATGGGACGGCTCCACCACCGACGAGTGGCAGTAGTGGAGATCCTT 422
 Qy 320 AspAlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGlu 339
 Db 423 GATGCCAATGACAAATGCTCCCATGTTTGACCCAGAGTACGAGGCCCATGTGCTGAG 482
 Qy 340 AsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSer 359
 Db 483 AATGCACTGGGCCATGAGGTGACAGGCTGACGGTCACTGATCTGCGACGCCCACTCA 542
 Qy 360 ProAlaThrArgAlaThrTyrLeuIleMetGlyValAspAspGlyAspHisPheThrIle 379
 Db 543 CCAGCGTGGCGTCCCACTACCTTATATCGGGCGTGACACGGGACCATTTTACATC 602
 Qy 380 ThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArg-LysGlyLeuAspPheGln 399

```

Db      603 ACCACCCACCTGAGAGCAACAGGCGCATCTTGACAAACAGGGAAGGGTTTGATTTTGA 662
QY      399 uAlaLysAsnGlnHisThrLeuTyrrValGluValThrAsnGluAlaProPheValLeuLys 419
Db      663 GGCACAAAACAGCAGCACACCTGTGATCTTGAGTGTACCAAGAGGCGCCCTTTTGCTGAA 722
QY      419 sLeuProThrSerThrAlaThrIleValValHis 430
Db      723 GCTCCCAACTCCACAGCCACCACCATAGTGTCCAC 756

RESULT 4
BM923882
LOCUS   BM923882
DEFINITION AGENCOURT_6709933 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5760016
5', mRNA sequence.
ACCESSION BM923882
VERSION   BM923882.1 GI:19374261
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1016)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-remail.nih.gov
          Tissue Procurement: Life Technologies, Inc.
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM12806 row: f column: 17
          High quality sequence stop: 703.

FEATURES
         Location/Qualifiers
         1..1016
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:5760016"
            /lab_host="DH10B"
            /clone_lib="NIH MGC 116"
            /note="Organ: pooled colon, kidney, stomach; Vector:
            pCMV-SPOK16; Site: 1: NotI; Site 2: EcoRV (destroyed); RNA
            source anonymous pool of 3 colons, age 26 yo male, 49 yo
            female, 71 yo male colon; 46 yo male kidney, and pool of 2
            stomachs, 62 yo male and 70 yo female. Library is
            oligo-dT primed and directionally cloned (EcoRV site is
            destroyed upon cloning). Average insert size 1.4 kb,
            insert size range 1-3 kb. Library is normalized and
            enriched for full-length clones and was constructed by C.
            Gruber (Invitrogen). Research Genetics tracking code
            023. Note: this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.:      1,68e-200      Length:      1016
Score:          235.00      Matches:      248
Percent Similarity: 99.60%      Conservative: 0
Best Local Similarity: 99.60%      Mismatches: 0
Query Match:     28.35%      Indels:      1
DB:             12          Gaps:         0

US-09-916-849A-1 (1-829) x BM923882 (1-1016)
QY      1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGluValCysTrpLeuGln 20
Db      45 ATGGGGCTCCCTCGTGAGACCTCTCGGCTCTCTCCCTCTCCAGGTTTGTGCTGCAGC 104

```

```

QY      21 CysAlaLysSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
Db      105 TGCGCGGCTCCGAGCGGTCGCCGGCGGTCTTCAGGGAGGCTGAAGTGAAGTCTGGAGCG 164
QY      41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db      165 GGAGCGCGGAGCAGGAGCGCGCCAGCGCTGGGAAAGTATTCATGGCTGCCCTGGG 224
QY      61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
Db      225 CAAGAGCCAGCTCTGTTAGCACTGATATGATGACTTCACTGTGCGGAATGGCGAGCA 284
QY      81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
Db      285 GTCCAGGAAGAAGGTCACTGAAGAAAGAAATCCATTAAGATCTTCCCATCAAAACGT 344
QY      101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
Db      345 ATCTTACGAAGACACAAGAGAGATTGGTGGTGTCTCCATATCTGTCCTGAAATGGC 404
QY      121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
Db      405 AAGGTCCTCCCTTCCCGCAGAGACTGAATCAGTCAAGTCTAATAAAGATGAGACACCAAG 464
QY      141 IlePheTyrrSerIleThrGlyProGlyValAlaAspSerProGluGlyValPheAlaVal 160
Db      465 ATTTCTACAGCATCACGGGCGCGGGGCGAGACAGCCCTCTGAGGGTGTCTTCGCTGA 524
QY      161 GluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluGluLeuAlaLys 180
Db      525 GAGAAGGAGACAGGCTGGTGTGTTGATTAAGCACTGGACCGGGAGAGATTGCCAAG 584
QY      181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyValaSerValGluAspProMetAsn 200
Db      585 TATGAGCTCTTTGGCCACGCTGTCTCAGAAATGGTCTCAGTGGAGGACCCCATGAAC 644
QY      201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThr-Ph 220
Db      645 ATCTCCATCATCTGTGACCGCAGATGACCAAGCCCAAGTTTACCAGGACACCCNTT 704
QY      220 eArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db      705 CCGAGGAGAGTGTCTTAGAGGAGTCTTACCAGGAGTACTTCTGTGATGAGGTGACAGCCAC 764
QY      240 rAspGluAspAlaIleTyrrThr 248
Db      765 AGATGAGATGATGCTTACACC 789

RESULT 5
CA489015
LOCUS   CA489015
DEFINITION AGENCOURT_10809410 MAPcL Homo sapiens cDNA clone IMAGE:6721249 5',
mRNA sequence.
ACCESSION CA489015
VERSION   CA489015.1 GI:24951806
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 875)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-remail.nih.gov
          Tissue Procurement: Kristi A. Eglund, Ira Pastan
          cDNA Library Preparation: Invitrogen Corp
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov

```

Plate: LLAM14282 row: b column: 01
High quality sequence stop: 647.

FEATURES

source

Location/Qualifiers

1. .875

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:6721249"

/cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,

hTERT-RME1, LNCaP"

/lab_host="EMDH10B"

/clone_lib="MAPcL"

/notes="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;

Subtracted with brain, liver, lung, kidney and muscle.

Directionally cloned. Priming method: oligo-dT. Average

insert size: 1800 bp. Library amplification: 26,000 fold.

Kristi A. Eglund, James J. Vincent, Robert Strausberg,

Bungkook Lee & Ira Pastan: Discovery of new breast

cancer genes encoding membrane and secreted proteins.

Manuscript submitted."

ORIGIN

Alignment Scores:

Pred. No.: 2.86e-194 Length: 875
Score: 228.00 Matches: 228
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 27.50% Indels: 0
DB: 14 Gaps: 0

US-09-916-849A-1 (1-829) x CA489015 (1-875)

576 ThrSerProPheGlnAlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluVal 595
576 ThrSerProPheGlnAlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluVal 595
35 ACTTCCCTTTCCAGGCCAGCTCAGATGATCTAGATCTACTGGAGCGGAGGTC 94
35 ACTTCCCTTTCCAGGCCAGCTCAGATGATCTAGATCTACTGGAGCGGAGGTC 94
596 AsnGluGlyAspThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyr 615
596 AsnGluGlyAspThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyr 615
95 AACGAGAGAGTGACACAGTGCTTGTCTCCCTGAAGAGTCTCTGAAGCAGGATACATAT 154
95 AACGAGAGAGTGACACAGTGCTTGTCTCCCTGAAGAGTCTCTGAAGCAGGATACATAT 154
616 AspValHisLeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAla 635
616 AspValHisLeuSerLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAla 635
155 GAGTGACACTTCTCTGTCTGACCTGCAACAAAGACAGCTGACGTGTATCAGGCC 214
155 GAGTGACACTTCTCTGTCTGACCTGCAACAAAGACAGCTGACGTGTATCAGGCC 214
636 ThrValCysAspCysHisGlyHisValGluThrCysProGlyProThrPheGlyGlyPhe 655
636 ThrValCysAspCysHisGlyHisValGluThrCysProGlyProThrPheGlyGlyPhe 655
215 ACTGTGTGACACTGCCATGCCATGTTCGAAACCTGCGCCCTGGACCTGGAAGGAGGTTTC 274
215 ACTGTGTGACACTGCCATGCCATGTTCGAAACCTGCGCCCTGGACCTGGAAGGAGGTTTC 274
656 IleLeuProValLeuGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeu 675
656 IleLeuProValLeuGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeu 675
275 ATCCTCCCTGTGTGGGGGTGTCCTGGGTCTGCTCTCTCTCTCTGTGTGTGTGTGTTTG 334
275 ATCCTCCCTGTGTGGGGGTGTCCTGGGTCTGCTCTCTCTCTCTGTGTGTGTGTGTTTG 334
676 LeuValArgLysLysArgLysIleLysGluProLeuLeuLeuProLysAspThrArg 695
676 LeuValArgLysLysArgLysIleLysGluProLeuLeuLeuProLysAspThrArg 695
335 TTGGTGAGAAAGACGAGATCAGAGAGCCCTCTCTACTCCAGAAAGATGACACCGCT 394
335 TTGGTGAGAAAGACGAGATCAGAGAGCCCTCTCTACTCCAGAAAGATGACACCGCT 394
696 AspAsnValPheTyrTyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIle 715
696 AspAsnValPheTyrTyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIle 715
395 GACAACGCTCTTCTACTATGCGAAGAGGGGGTGGCGAAGAGAGCAGACACTATGACATC 454
395 GACAACGCTCTTCTACTATGCGAAGAGGGGGTGGCGAAGAGAGCAGACACTATGACATC 454
716 ThrGlnLeuHisArgGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAla 735
716 ThrGlnLeuHisArgGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAla 735
455 ACCAGCTCACCGAGGTCTGGAGGCCAGGCCGGAGGTGTTCTCGCAATGAGTGGCA 514
455 ACCAGCTCACCGAGGTCTGGAGGCCAGGCCGGAGGTGTTCTCGCAATGAGTGGCA 514
736 ProThrIleIleProThrProMetTyrArgProArgProAlaAsnProAspGluIleGly 755
736 ProThrIleIleProThrProMetTyrArgProArgProAlaAsnProAspGluIleGly 755
515 CCAACATCATCCGACACCATGTATCCGTCTCTCGGCCAGCCCAACCCAGATGAAATCGGC 574
515 CCAACATCATCCGACACCATGTATCCGTCTCTCGGCCAGCCCAACCCAGATGAAATCGGC 574
756 AsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAspProThrAlaProThrAsp 775
756 AsnPheIleIleGluAsnLeuLysAlaAlaAsnThrAspProThrAlaProThrAsp 775
575 AACTTTATAATGAGAACTGAAGGGGGCTTAACACAGACCCCAACAGCCCGCCCTACGAC 634
575 AACTTTATAATGAGAACTGAAGGGGGCTTAACACAGACCCCAACAGCCCGCCCTACGAC 634
776 ThrLeuLeuValPheAspTyrGluGlySerGlySerAspAlaAlaSerLeuSerLeu 795
776 ThrLeuLeuValPheAspTyrGluGlySerGlySerAspAlaAlaSerLeuSerLeu 795

Db 635 ACCCTCTGTGTCTGACTATGAGGCGGCTCCGACCGCGTCCCTGAGCTCCCTC 694
Qy 796 ThrSerSerAlaSerAspGlnAsp 803
Db 695 ACCTCTCTCGGCTCCGACCAAGAC 718
RESULT 6
BM545620 1122 bp mRNA linear EST 20-FEB-2002
LOCUS AGENCOURT 6500297 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5587889
DEFINITION 5', mRNA sequence.
ACCESSION BM545620
VERSION BM545620.1 GI:18777871
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1122)
NIH-MGC http://mgi.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: c9apbs-x@mail.nih.gov
Tissue Procurement: Invitrogen
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12358 row: b column: 18
High quality sequence start: 108
High quality sequence stop: 742.
Location/Qualifiers
1. .1122
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5587889"
/lab_host="DH10B"
/clone_lib="NIH_MGC_125"
/notes="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
Site 1: EcoRV (destroyed); Site 2: NotI; RNA source pool
of three ovaries, from females ranging in age from 38 to
49 yo. Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 2.1 kb, insert size range 1-3.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 036."

FEATURES

source

ORIGIN

Alignment Scores:

Pred. No.: 3.7e-194 Length: 1122
Score: 228.00 Matches: 244
Percent Similarity: 99.59% Conservative: 0
Best Local Similarity: 99.59% Mismatches: 1
Query Match: 27.50% Indels: 1
DB: 12 Gaps: 0

US-09-916-849A-1 (1-829) x BM545620 (1-1122)

Qy 585 AspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAspThrValValLeu 604
Db 120 GATGACTCAGACATCTACTCGAGCGGAGGTCAACGAGGAAGTGACCA-GTGGTCTTG 178
Qy 605 SerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSerLeuSerAspHis 624
Db 179 TCCCTGAGAGAGTTCCTGAAGCAGGATACATATGACGTGCACCTTCTCTGTCTGACCAT 238
Qy 625 GlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCysHisGlyHisVal 644

```

Db 239 GGCACAAAGAGAGCGTGCAGTGCATGAGGCGACCTGTGTGCGATGCGATGGCCATGTC 298
QY 645 GluThrCysProGlyProTyrPheLeuValLeuProValLeuGlyAlaValLeu 664
Db 299 GAAACCTGCGCTGACCCCTGGAAGGAGGTTTCATCTCTCCCTGCTGCTGGGCTGCTG 358
QY 665 AlaLeuLeuPheLeuLeuValLeuLeuLeuValLeuValLeuValLeuValLeuValLeu 684
Db 359 GCTCTGCTGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 418
QY 685 GluProLeuLeuLeuProGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 704
Db 419 GAGCCCTCTCTACTCCAGAGATGACACCCGTCGACAGCTCTTCTACTATGGCGAAGAG 478
QY 705 GlyGlyGlyGluGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 724
Db 479 GGGGGTGGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 538
QY 725 ArgProGluValValLeuValLeuValLeuValLeuValLeuValLeuValLeuValLeu 744
Db 539 AGCCCGAGTGGTCTCTCGCATGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 598
QY 745 ArgProArgProAlaAsnProAspGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 764
Db 599 CGTCTCGGCCAGCCAGCCAGATGAATCGGCACTTTATTAATGAGAACCTGGAAGCG 658
QY 765 AlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPheAspTyrGluGly 784
Db 659 GCTAACACAGACCCACAGCCCGCCCTACGACACCCCTCTGGTGTGCTGCTGCTGCTGCT 718
QY 785 SerGlySerAspAlaAlaSerLeuSerLeuSerLeuSerLeuSerLeuSerLeuSerLeu 804
Db 719 AGCGGCTCGAGCGCGCTCCCTGAGTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 778
QY 805 AspTyrAspTyrLeuAsnGluTrpGlySerArgPheLeuLeuLeuLeuLeuLeuLeuLeu 824
Db 779 GATTACGATTATCTGAACAGTGGGCGAGCCGCTTCAAGAGACTGGCAGACATGACGGT 838
QY 825 GlyGlyGluAspAsp 829
Db 839 GCGGGGAGGAGCGAC 853

RESULT 7
BG403092
LOCUS
DEFINITION BG403092 769 bp mRNA linear EST 12-MAR-2001
mRNA sequence.
ACCESSION BG403092
VERSION BG403092.1 GI:13296540
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 769)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL0432 row: f column: 03
High quality sequence stop: 714.
Location/Qualifiers
1..769
/organism="Homo sapiens"

```

FEATURES
source

ORIGIN

```

Alignment Scores:
Pred. No.: 1.99e-193 Length: 769
Score: 227.00 Matches: 227
Percent Similarity: 100.00% Conservations: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 27.38% Indels: 0
DB: 12 Gaps: 0

US-09-916-849A-1 (1-829) x BG403092 (1-769)

QY 508 PheValArgAsnAsnLeuTyrGluValMetValLeuAlaMetAspAsnGlySerProPro 527
Db 2 TTTCTGAGGAACAACATCTTATGAAGTCATGCTTGGCCATGGACAATGGAAGCCCTCCC 61
QY 528 ThrThrGlyThrGlyThrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 547
Db 62 ACCACTGGCAGCGGAACCTTCTGTCTAACACTGATGATGTCACAGACCATGGCCCGAGT 121
QY 548 ProGluProArgGlnIleThrIleCysAsnGlnSerProValArgHisValLeuAsnIle 567
Db 122 CTTGAGCCCGCTCAGATCACCATCTGCAACCAAGCCCTGTGCGCCACGTCGTGAACATC 181
QY 568 ThrAspLysAspLeuSerProHisThrSerProPheGlnAlaGlnLeuThrAspAspSer 587
Db 182 ACGGACAAGACCTGTCTCCGCCACCTCCCTTTTCCAGGCCCCAGCTTCACAGATGACTCA 241
QY 588 AspIleTyrTrpThrAlaGluValAsnGluGluGlyAspThrValValLeuSerLeuLys 607
Db 242 GACATCTACTGGACGGCAGAGGTCAACAGAGAGGTGACACAGTGTCTTGTCCCTGAAG 301
QY 608 LysPheLeuLysGlnAspThrTyrAspValHisLeuSerLeuSerAspHisGlyAsnLys 627
Db 302 AAGTTCTCTGAAGCAGGATACATATGACGTGCACCTTTCTCTGTGACCATGGCAACAAA 361
QY 628 GluGlnLeuThrValIleArgAlaThrValCysAspCysHisGlyHisValGluThrCys 647
Db 362 GAGCAGCTGACGGTATCAGGGCCACTGTGTGCGACTGCGCATGGCCATGTCGAAACCTGC 421
QY 648 ProGlyProTyrLysGlyGlyPheIleLeuProValLeuGlyAlaValLeuAlaLeuLeu 667
Db 422 CTTGAGCCCTGGAAAGGAGGTTTCATCTCTCCCTGTGCTGGGGCTGTCTGTGCTGCTG 481
QY 668 PheLeuLeuLeuValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLysLysLysLys 687
Db 482 TTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 541
QY 688 LeuLeuProGluLeuAspThrArgAspAsnValPheTyrTyrGlyGluGluGlyGly 707
Db 542 CTACTCTCCAGAGATGACACCCGTCGACACGCTTCTACTATGGCGAAGGGGGGTGGC 601
QY 708 GluGluAspGlnAspTyrAspIleThrGlnLeuHisArgGlyLeuGluAlaArgProGlu 727
Db 602 GAAGAGACCGAGGACTATGACATCACCCAGCTCCAGAGTCTGGAGGCGAGGCGGAG 661
QY 728 ValValLeuArgAsnAspVal 734
Db 662 GTGGTTCTCCGCAATGACGTG 682

```

RESULT 8
CA455046

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4525826"
/tissue_type="transitional cell papilloma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_93"
/note="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dr primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

```

```

LOCUS       CA455046                887 bp    mRNA    linear    EST 12-NOV-2002
DEFINITION  AGENCOURT 10714428 MAPcL Homo sapiens cDNA clone IMAGE:6722552 5',
            mRNA sequence.
ACCESSION   CA455046
VERSION     CA455046.1  GI:24905371
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 887)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabbs@mail.nih.gov
            Tissue Procurement: Kristi A. Eglund, Ira Pastan
            CDNA Library Preparation: Invitrogen Corp
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14285 row: h column: 08
            High quality sequence stop: 594.
FEATURES             Location/Qualifiers
                     1..887
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:6722552"
                        /cell_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,
                        HER2-1, LNCaP"
                        /lab_host="EMD108"
                        /clone_lib="MAPcL"
                        /notes="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;
                        Subtracted with brain, liver, lung, kidney and muscle.
                        Directionally cloned. Priming method: oligo-dT. Average
                        insert size: 1800 bp. Library amplification: 26,000 fold.
                        Kristi A. Eglund, James J. Vincent, Robert Strausberg,
                        Bungkok Lee & Ira Pastan. Discovery of new breast
                        cancer genes encoding membrane and secreted proteins.
                        Manuscript submitted."
ORIGIN
Alignment Scores: 1.15e-190 Length: 887
Pred. No.: 224.00 Matches: 251
Score: 99.21% Conservative: 0
Percent Similarity: 99.21% Mismatches: 1
Best Local Similarity: 27.02% Indels: 2
Query Match: 27.02% Gaps: 0
DB: 14

US-09-916-849A-1 (1-829) x CA455046 (1-887)

QY 287 AspArgGluLysValProGluTyrThrLeuThrIleGlnAlaThrAspMetAspGlyAsp 306
Db 3 GACCGGMAAAAGTCCTCGAGTACACTGACCTCAGCCAGGCCACACATGATGGGAC 62
QY 307 GlySerThrThrAlaValAlaValGluLeuAlaAspAsnAlaPro 326
Db 63 GGTTCACACACACCGGAGGAGTAGTGAGATCCTTGATGCCAATGCAATGCTCCC 122
QY 327 MetPheAspProGlnLysTyrGluAlaHisValProGluAsnAlaValGlyHisGluVal 346
Db 123 ATCTTTGACCCCGAGAGTAGAGGCCCATGTGCTGAGATGATGAGTGGCCATGAGGTG 182
QY 347 GlnArgLeuThrValThrAspLeuAspAlaProAsnSerProAlaThrArgAlaThrTyr 366
Db 183 CAGAGGCTGACGCTCACTGATCTGGACGCCCACTCAGCGCGGGTGGTCCACCTAC 242
QY 367 LeuIleVetGlyGlyAspGlyAspHisPheThrIleThrHisProGluSerAsn 386

```

```

Db 243 CTTATCATGGCGGTGACGACGGGACCATTTTACCATCACCACCACCTCGAGACAAC 302
QY 387 GlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAlaLysAsnGlnHisThrLeu 406
Db 303 CAGGGCATCTGACACACAGGAGGTTTGATTTTGGAGCCAAAACACAGCACACCTTG 362
QY 407 TyrValGluValThrAsnGluAlaProPheValLeuLysLeuProThrSerThrAlaThr 426
Db 363 TACCTTGAAGTGAACACAGGAGGCTTTGTGCTGAAGCTCCAACTCCACAGGCACC 422
QY 427 IleValValHisValGluAspValAsnGluAlaProValPheValProProSerLysVal 446
Db 423 ATAGTGTTCACGTGGAGGATGTGAATGAGGCACCTGTGTTTCCACCTCCAAAGTC 482
QY 447 ValGluValGlnGluGlyIleProThrGlyGluProValCysValTyrThrAlaGluAsp 466
Db 483 GTTGAGGTCCAGGAGGACATCCCACTGGGGAGCCTGTGTGTCTACACTGCAGAGAC 542
QY 467 ProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArgAspProAlaGlyTyrLeu 486
Db 543 CTTGACAGGAGATCAAAAGATCAGCTACCGCATCTGAGAGACCCAGCAGGTTGGCTA 602
QY 487 AlaMetAspProAspSerGlyGlnValThrAlaValGlyThrLeuAspArgGluAspGlu 506
Db 603 GCCATGGACCCAGACAGTGGGAGGTACAGCTGTGGCACCCTCGACCTGAGGATGAG 662
QY 507 GlnPheValArg-AsnAsnIleTyrGluValMetValLeuAlaMetAspAsnGlySerPr 526
Db 663 CAGTTTGTGAGGAACAC-ATCTATGAAGTCATGTGCTTGGCCATGACAAATGAAGCCC 721
QY 526 cProThrThrGlyThrGlyThrLeuLeuLeuThrLeu 538
Db 722 TCCACCACTGGCAGCGGAACCTTCTGCTAACACTG 758

RESULT 9
BU543746
LOCUS       BU543746                865 bp    mRNA    linear    EST 13-SEP-2002
DEFINITION  AGENCOURT 10324520 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6575911
            5', mRNA sequence.
ACCESSION   BU543746
VERSION     BU543746.1  GI:22854229
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 865)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabbs@mail.nih.gov
            Tissue Procurement: DCTD/DTF
            CDNA Library Preparation: Rubin Laboratory
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LICM2773 row: j column: 07
            High quality sequence stop: 654.
FEATURES             Location/Qualifiers
                     1..865
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:6575911"
                        /tissue_type="carcinoma, cell line"
                        /lab_host="DH10B (phage-resistant)"
                        /clone_lib="NIH_MGC_40"
                        /notes="Organ: prostate; Vector: pOTB7; Site 1: XhoI;
                        Site 2: EcoRI; cDNA made by oligo-dT priming.
                        Directionally cloned into EcoRI/XhoI sites using the

```


following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:

Pred. No.: 5,63e-188 Length: 865
Score: 221.00 Matches: 221
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 26.66% Indels: 0
DB: 13 Gaps: 0

US-09-916-849A-1 (1-829) x BU543746 (1-865)

QY 178 IleAlaValThrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAsp 197
Db 3 ATTGCCAAGTATGAGCTCTTTGGCCACGCTGTGTACAGAAATGGTCCATCATGAGGAC 62
QY 198 ProMetAsnIleSerIleValThrAspGlnAsnAspHisGlyProLysPheThrGln 217
Db 63 CCATCAACATCTCCATCATCTGTCGACCGACCAAGATGACCAAGCCCAAGTTTACCCAG 122
QY 218 AspThrPheArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnVal 237
Db 123 GACACCTTCGAGGGAGTCTCTAGAGGAGTCTTACCAGAGTCTCTGTATGAGGTG 182
QY 238 ThrAlaThrAspGluAspAlaIleThrThrTyAsnGlyValValAlaValSerIle 257
Db 183 ACAGCCACGGATGAGATGATCCATCTACATCTACATGAGGGTGTCTTACTCCATC 242
QY 258 HisSerGlnGluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGly 277
Db 243 CATAGCAAGAACCAAGAGCCACACGACCTCATGTTCCACATTCACCGAGCACAGGC 302
QY 278 ThrIleSerValIleSerGlyLeuAspArgGlyValProGluThrThrLeuThr 297
Db 303 ACATCAGCGTATCTCCAGTGCCTTGGACCGGAAAGTCCCTGAGTACATCTGACC 362
QY 298 IleGlnAlaThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGlu 317
Db 363 ATCCAGGCACACATGATGCGGACGCTCCACACCGACGAGTGCAGTGTAGGAG 422
QY 318 IleLeuAspAlaAsnAspAsnAlaProMetPheAspProGlnLysTyGluAlaHisVal 337
Db 423 ATCCCTTGATGCCAATGACATGCTCCCATGTTTGAACCCCAAGATGACAGGCCCATGTG 482
QY 338 ProGluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaPro 357
Db 483 CTTGAGATGCGATGGCCATGAGTGTGACAGGCTGACGCTCACTGATCTGACGCCCCC 542
QY 358 AsnSerProAlaThrArgAlaThrTyLeuIleMetGlyGlyAspAspHisPhe 377
Db 543 AACTCACCAGCGTGGCGTCCACCTACCTTATCATGGGCGGTGACGAGCGGACCATTT 602
QY 378 ThrIleThrThrHisProGluSerAsnGlnGlyLeuLeuThrThrArgGlyGlyLeuAsp 397
Db 603 ACATCACCACCCACCTGAGAGCAACACAGGCGATCTTGACAAACACGAGGAGTTTGGAT 662
QY 398 Phe 398
Db 663 TTT 665

RESULT 10

BM715802

LOCUS

DEFINITION

BM715802

ACCESSION

VERSION

KEYWORDS

BM715802 703 bp mRNA linear EST 28-FEB-2002
UI-E-EJO-ahj-m-06-0-UI.r2 UI-E-EJO Homo sapiens cDNA clone
UI-E-EJO-ahj-m-06-0-UI 5', mRNA sequence.
BM715802
BM715802.1 GI:19029060
EST.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (Bases 1 to 703)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9585

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

Seq primer: M13 Reverse.

Location/Qualifiers

1..703

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-E-EJO-ahj-m-06-0-UI"

/tissue_type="fetal eyes, lens, eye anterior segment,

optic nerve, retina, Retina Foveal and Macular, RPE and

Choroid"

/dev_stage="fetal and adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone_lib="UI-E-EJO"

/notes="Organ: eye; Vector: pT73-Pac (Pharmacia) with a

modified polylinker; Site: 1: EcoR I; Site 2: Not I;

UI-E-EJO is a subtracted cDNA library constructed

according to Bonaldo, Lennon and Soares, Genome Research,

6:791-806, 1996. First strand cDNA synthesis was primed

with an oligo-dT primer containing a Not I site. Double

stranded cDNA was ligated to an EcoR I adaptor, digested

with Not I, and cloned directionally into pT73-Pac

vector. The oligonucleotide used to prime the synthesis of

first-strand cDNA contains a library tag sequence that is

located between the Not I site and the (dT)18 tail. The

sequence tags for this library are: fetal eyes,

AGATCAAGA; lens, CGATTAGCGA; eye anterior segment,

AATCCCGCAT; optic nerve, CCATTAGTG; retina, CCGCG;

Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This

library was created for the program, Gene Discovery in the

Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Alignment Scores:

Pred. No.: 9.05e-182 Length: 703
Score: 214.00 Matches: 227
Percent Similarity: 99.56% Conservative: 0
Best Local Similarity: 99.56% Mismatches: 0
Query Match: 25.81% Indels: 1
DB: 12 Gaps: 0

US-09-916-849A-1 (1-829) x BM715802 (1-703)

QY

Db

QY

Db

336 HisValProGluAsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAsp 355

6 CATGTGCTGAGATGTCAGTGGGCGCATGAGTGCAGAGGCTGACGCTCATCTGTCGAC 65

356 AlaProAsnSerProAlaThrTcTcGAlaThrTyLeuIleMetGlyGlyAspGlyAsp 375

66 GCCCCCAACTCACAGCGTGGCGTGCACCTATCATGGCGGTGACGAGGGGAC 125

376 HisPheThrIleThrThrHisPro-GluSerAsnGlnGlyIleLeuThrThrArgLysGI 395
126 CATTTTACCATCACCAACCTTGAGAGCAACAGGCGATCTGTGCAACACAGAGGG 185
395 YLeuAspPheGluAlaLysAsnGlnHisThrLeuThrValGluValThrAsnGluAlaPr 415
186 TTGGATTGAGCCCAAAACAGACACACCTGTACGTGTAGTGAAGTGAACACAGGCC 245
415 oPheValLeuLysLeuProThrSerThrAlaThrIleValValHisValGluAspValAs 435
246 TTTTGTGCTGAAGCTCCCAACCTCCACAGCCACCATAGTGTGCTCCAGTGGAGATGTGAA 305
435 nGluAlaProValPheValProSerLysValValGluValGluGlyIleProTh 455
306 TGAGGCACCTGTGTTTGTCCCAACCTCCAAAGTCGTTGAGGTCCAGGAGGATCCCCAC 365
455 rGlyGluProValCysValThrThrAlaGluAspProAspLysGluAsnGlnLysIleSe 475
366 TGGGAGCGCTGTGTGTCTACACTGCAGAGACCCCTGCAAGGAGATCAAAAGATCAG 425
475 rThrArgIleLeuArgAspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGluVa 495
426 CTACCCGATCTCTGAGAGACCCAGAGCGGTGGCTAGCCATGGACCCAGACAGTGGCAGGT 485
495 lThrAlaValGlyThrLeuAspArgGluAspGluGlnPheValArgAsnAsnIleThrGI 515
486 CACAGCTGTGGGACCCCTCGACCGTGAGTGAAGTGTGAGGAGAACACATCATGA 545
515 uValMetValLeuAlaMetAspAsnGlySerProThrThrGlyThrGlyThrLeuLe 535
546 AGTCATGCTTGTGGCATGTGACCAATGAAGCCCTCCACCATCTGGCAGCGGAACCTTCT 605
535 uLeuThrLeuLeuAspValAsnAspHisGlyProValProGluProArgGlnIleThrII 555
606 GCTAACACTGATTGATGTCAATGACCATGGCCAGTCCCTGAGCCCGTCAGATCACCAT 665
555 eCysAsnGlnSerProValArg 562
666 CTGCAACCAAGCCCTGTGGGT 687

RESULT 11
BUI68948 1006 bp mRNA linear EST 04-SEP-2002
LOCUS AGENCOURT_8110724 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6266745
DEFINITION 5', mRNA sequence.
ACCESSION BUI68948.1 GI:22682932
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE Tissue Procurement: DCTD/DTP
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1006)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-i@mail.nih.gov
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LILNL at:
<http://image.llnl.gov>
Plate: LLCM2435 row: p column: 10
High quality sequence stop: 607.
Location/Qualifiers
1..1006
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

/clone="IMAGE:6266745"
/tissue type="melanotic melanoma, cell line"
/lab host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_112"
/notes="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN

Alignment Scores: 1,31e-181 Length: 1006
Pred. No.: 214.00 Matches: 214
Score: 214.00
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 25.81% Indels: 0
DB: 13 Gaps: 0
US-09-916-849A-1 (1-829) x BUI68948 (1-1006)
QY 320 AsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSer 359
Db 2 GATGCCAATGCAATGCTCCCATGTTGACCCCAAGTACGAGGCCCATGTGCTTGTGAG 61
QY 340 AsnAlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSer 359
Db 62 AATGCATGTGGCCATGAGTGCAGAGGCTGACGTCATCATCTGACGCCCCCACTCA 121
QY 360 ProAlaTrpArgAlaThrThrLeuLeuMetGlyValAspAspGlyAspHisPheThrIle 379
Db 122 CCAGCGTGGGTGGCCACTTACCTTATCATGGCGGTGACACGGGACCACTTTTACATC 181
QY 380 ThrThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGlu 399
Db 182 ACCACCCACCTTGAGACCAACAGGCGCATCTGCAACACAGGAGGGTTTGGATTTTGA 241
QY 400 AlalysAsnGlnHisThrLeuThrValGluValThrAsnGluAlaProPheValLeuLys 419
Db 242 GCCAAAAACACGACACACCTGTACGTTGAAGTACCAACACGAGGCCCTTTTGTGCTGA 301
QY 420 LeuProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProVal 439
Db 302 CTCCCAACCTCCACAGCACCATAGTGTCCACGTGGAGATGTGATGAGGACCTGTG 361
QY 440 PheValProSerLysValValGluValGlnGlyIleProThrGlyGluProVal 459
Db 362 TTTGTCCACCTCCAAAGTCGTTGAGGTCCAGAGGCGCATCCCACTGGGAGCGCTGTG 421
QY 460 CysValThrThrAlaGluAspProAspLysGluAsnGlnLysIleSerThrArgIleLeu 479
Db 422 TGTGTCTACACTGCAGAGACCCCTGCAAGGAGATCAAAAGATCAGCTTACCGCATCTG 481
QY 480 ArgAspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGly 499
Db 482 AGAGACCCAGCAGGGTGGCTAGCCATGACCCACAGTGGCGAGGTCCACAGCTGTGGGC 541
QY 500 ThrLeuAspArgGluAspGluGlnPheValArgAsnAsnIleThrGluValMetValLeu 519
Db 542 ACCCTCGACCGTGGAGTGAAGTGTGAGGAGAACCAATCTATGAGTGAAGTCAATGCTTG 601
QY 520 AlaMetAspAsnGlySerProThrThrGlyThrGlyThr 533
Db 602 GCCATGGAATGAAGCCCTCCCACTGGCACGGGAACC 643

RESULT 12
CB240552/c
LOCUS CB240552
DEFINITION UI-CF-PNO-afv-b-02-0-UI.s1 UI-CF-PNO Homo sapiens cDNA clone
UI-CF-PNO-afv-b-02-0-UI.3', mRNA sequence.
745 bp mRNA linear EST 12-FEB-2003

```
ACCESSION      CB240552
VERSION        CB240552.1
KEYWORDS       GI:28362196
SOURCE         EST.
ORGANISM       Homo sapiens (human)

REFERENCE
AUTHORS        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL        1 (bases 1 to 745)
MEDLINE        Bonaldo,M.F., Lennon,G. and Soares,M.B.
PURNED         Normalization and subtraction: two approaches to facilitate gene
COMMENT         discovery
                Genome Res. 6 (9), 791-806 (1996)
                97044477
                8889548
                Contact: McCray, PB
                McCray Lab
                University of Iowa
                2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
                Tel: 319 356 4866
                Fax: 319 356 7171
                Email: paul-mccray@uiowa.edu
                Tissue Procurement: Dr. M. J. Welsh, University of Iowa
                cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
                cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
                DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
                Clone Distribution: Researchers may obtain clones from Research
                Genetics (www.resgen.com) or from Open Biosystems
                (www.openbiosystems.com).
                Seq primer: M13 FORWARD
                POLYA=yes.

FEATURES
    source
    1..745
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="UI-CF-FNO-afv-b-02-0-UI"
        /tissue_type="Human Lung Epithelial cells"
        /lab_host="PH10B (Life Technologies) (T1 phage resistant)"
        /clone_lib="UI-CF-FNO"
        /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
        modified polylinker; Site 1: EcoR I; Site 2: Not I;
        UI-CF-FNO is a subtracted cDNA library derived from two
        normalized Human lung epithelial cell libraries (EN1 and
        DU1) The library was subtracted according to according to
        Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
        1996. For additional information, contact:
        bento-soares@uiowa.edu
        TAG_TISSUE=Human Lung Epithelial Cell Lines untreated LPS
        6hr to LPS 24h
        TAG_LIB=UI-CF-FNO
        TAG_SEQ=CTGCTCAGGT"

ORIGIN
Alignment Scores:
Pred. No.:      7,638-181      Length:      745
Score:          213.00      Matches:      213
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    25.69%      Indels:      0
DB:            14      Gaps:      0

US-09-916-849A-1 (1-829) x CB240552 (1-745)

QY      53      LysValPheMetGlyCysProGlyGlnGluProAlaLeuPheSerThrAspAsnAspAsp 72
DB      659    AAAGTATTCTGGCTGCCCTGGCAGAGAGCCAGCTCTGTTAGCACTGATAATGATGAC 600
QY      73      PheThrValArgAsnGlyGluThrValGlnGluArgArgSerLeuLysGluArgAsnPro 92
DB      599    TTCACGTGCGGAATGGCAGACAGTCACAGGAAGAGGTCACTGAGGAAGAAGGAATCCA 540
QY      93      LeuLysIlePheProSerLysArgIleLeuArgArgHisLysArgAspTrpValValAla 112
```



```

Db      404  AAGGGTCCCTTCCCCAGAGACGTAATCAGCTCAAGTCTAATAAAGATAGACACCAAG 463
QY      141  IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
Db      464  ATTTCTACAGATACAGGGCGGGGCGAGACAGCCCCCTTGGGTGTCTTCGTGTA 523
QY      161  GluLysGluThrGlyTTrpLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 180
Db      524  GAGAAGGAGACAGGCTGGTGTGTGTGAATAAGCCACTGGACCGGGAGAGATTGCCAAG 583
QY      181  TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db      584  TATGACTCTTTGGCCACGCTGTGTGAGAGATGGTGCCTCAGTGGAGACCCCATGAC 643
QY      201  IleSerIleValThrAspGlnAsnAspHis 211
Db      644  ATCTCATCATCGTGACCGACCAAGATGACCAC 676

RESULT 15
BM804940
LOCUS   BM804940
DEFINITION AGENCOURT_6495104 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5725677
5', mRNA sequence.
ACCESSION BM804940
VERSION   BM804940.1 GI:19121763
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 999)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgabbs@mail.nih.gov
            Tissue Procurement: Invitrogen
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12716 row: 0 column: 22
            High quality sequence start: 18
            High quality sequence stop: 680.
FEATURES
source    Location/Qualifiers
            1..999
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:5725677"
                /lab_host="DH10B"
                /clone_lib="NIH_MGC_125"
                /notes="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
                Site: 1: EcoRV (destroyed); Site 2: NotI; RNA source pool
                of three ovaries, from females ranging in age from 38 to
                49 yo. Library is oligo-dT primed and directionally cloned
                (EcoRV site is destroyed upon cloning). Average insert
                size 2.1 kb, insert size range 1-3.5 kb. Library is
                normalized and enriched for full-length clones and was
                constructed by C. Gruber (Invitrogen). Research Genetics
                tracking code 036."

```

ORIGIN

```

Alignment Scores:
Pred. No.:      3,27e-176      Length:      999
Score:          208.00         Matches:    208
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:    25.09%         Indels:     0
DB:              12           Gaps:      0

```

US-09-916-849A-1 (1-829) x BM804940 (1-999)

```

QY      54  ValPheMetGlyCysProGlyGlnGluProAlaLeuPheSerThrAspAsnAspPhe 73
Db      118  GTATTCTATGGGCTGCCCTGGGCAAGAGCCAGCTCTGTTTAGCACTGATAATGATGACTTC 177
QY      74  ThrValArgAsnGlyGluThrValGlnGluArgSerLeuLysGluArgAsnProLeu 93
Db      178  ACTGTGGGAAATGGCGAGACAGTCCAGGAAGAAGAGTCACTGAAGGAAAGAAATCCATTG 237
QY      94  LysIlePheProSerLysArgIleLeuArgArgHisLysArgAspTrpValValAlaPro 113
Db      238  AAGATCTTCCCATCCAAACGATATCTTACGAGACACAGAGAGATTGGTGGTGGTCTCCA 297
QY      114  IleSerValProGluAsnGlyLysGlyProPheProGlnArgLeuAsnGlnLeuLysSer 133
Db      298  ATATCTGTCCCTGAAATGGCAAGGGTCCCTTCCCCAGAGACTGAATCAGCTCAAGTCT 357
QY      134  AsnLysAspArgAspThrLysIlePheTyrSerIleThrGlyProGlyAlaAspSerPro 153
Db      358  AATAAGATAGACACCAAGATTTTCTACAGCATCACGGGGCCCGGGGCGAGACAGCCCC 417
QY      154  ProGluGlyValPheAlaValGluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeu 173
Db      418  CCTGAGGGTGTCTTCGCTGTAGAGAAGGAGACAGGCTGGTGTGTGTGAATAAGCCACTG 477
QY      174  AspArgGluGluIleAlaLysTyrGluLeuPheGlyHisAlaValSerGluAsnGlyAla 193
Db      478  GACCGGGAGGAGATGGCCAAAGTATGAGCTCTTTGGCCACGCTGTGTCAGAGAATGGTGCC 537
QY      194  SerValGluAspProMetAsnIleSerIleValThrAspGlnAsnAspHisLysPro 213
Db      538  TCAGTGGAGGACCCCATGAACATCTCCATCATCTGACCGACCAAGATGACCACCAAGCCC 597
QY      214  LysPheThrGlnAspThrPheArgGlySerValLeuGluGlyValLeuProGlyThrSer 233
Db      598  AAGTTTACCAGACACCTTCCGAGGGAGTGTCTTAGAGGGGAGTCTTACCAAGTACTTCT 657
QY      234  ValMetGlnValThrAlaThrAspGluAspAlaIleTyrThrTyrAsnGlyValVal 253
Db      658  GTGATGCAAGTGACAGCCACCGATGAGATGATGCCATCTACACCTACATGGGGTGGTT 717
QY      254  AlaTyrSerIleHisSerGlnGlu 261
Db      718  GCTTACTCATCCATAGCCAAAGAA 741

```

Search completed: September 23, 2004, 07:38:32
Job time : 5105 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 23, 2004, 01:28:20 / Search time 8137 Seconds
(without alignments)
4415.801 Million cell updates/sec

Title: US-09-916-849A-1

Perfect score: 829

Sequence: 1 MGLPRGPIASLLLLQVCWLQ.....NEWGRFKKLADMYGGGEDD 829

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6934743

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cpn2.1/USPTO.spool.h/US09916849/runat.21092004.145835.7317/app_query.fasta.1.967
-DB=GenEmbl -QWTF=fastap -SUFFIX=oligo.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=quality -THR WIN=1 -ALIGN=15 -MODE=LOCAL -OUTFWT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09916849 @CGN 1.1.5239 @runat.21092004.145835.7317 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=60
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.ste.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.ste.*
28: em.un.*

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pin.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	829	100.0	3171	6	AR270785 Sequence
2	829	100.0	3171	6	AX411115 Sequence
3	829	100.0	3171	6	AX823232 Sequence
4	829	100.0	3171	9	X63629 H.sapiens m
5	715	86.2	3219	6	AX382118 Sequence
6	627	75.6	3686	9	BC041846 Homo sapi
7	102	12.3	306	6	AX210689 Sequence
8	80	9.7	184945	9	AC099314 Homo sapi
9	80	9.7	186839	30	AL139227 Human chr
10	80	9.7	186911	9	AC126773 Homo sapi
11	72	8.7	54207	2	AF002024 Homo sapi
12	61	7.4	2106	4	BTCDHP
13	54	6.5	255	6	AX210739 Sequence
14	51	6.2	2993	10	BC052189
15	51	6.2	3187	10	MCADHP
16	37	4.5	316	10	AF177683 Rattus no
17	28	3.4	179743	2	AC144675 Rattus no
18	28	3.4	190915	2	AC132132 Mus muscu
19	28	3.4	219657	2	AC129611 Rattus no
20	28	3.4	281505	2	AC120695 Rattus no
21	27	3.3	435	4	AF033826 Sus scrof
22	27	3.3	179743	2	AC144675 Rattus no
23	23	2.8	241282	2	AC094652 Rattus no
24	21	2.5	522	6	AX393688 Sequence
25	17	2.1	1055	5	CHKCAMA16
26	17	2.1	2436	5	GGCADHB
27	17	2.1	3545	5	CHKLCAMR
28	17	2.1	4434	5	CHKCAM
29	16	1.9	1684	5	XLBCAD
30	16	1.9	2310	6	AX598677 Sequence
31	16	1.9	2699	5	XLEPCAD
32	16	1.9	3450	5	XLU04707
33	16	1.9	3750	5	XLXBCADH
34	16	1.9	5310	6	AX767362 Sequence
35	15	1.8	164	10	MMUVEX15
36	15	1.8	168	5	CHKCAMA15
37	15	1.8	204	10	RNECADHER
38	15	1.8	313	10	AF177680 Rattus no
39	15	1.8	839	11	BV049696
40	15	1.8	923	9	HSPCADHGN
41	15	1.8	1902	10	MMUVEX16
42	15	1.8	2173	5	XLMECAD
43	15	1.8	2269	9	HSECAD16
44	15	1.8	2269	9	HUMCDH1014
45	15	1.8	2486	10	MMUVOM

ALIGNMENTS

RESULT 1

AR270785
 LOCUS AR270785 3171 bp DNA linear PAT 10-APR-2003
 DEFINITION Sequence 1348 from patent US 6500938.
 ACCESSION AR270785
 VERSION AR270785.1 GI:29702019
 KEYWORDS Unknown.
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 3171)
 AUTHORS Au-Young, J. and Seilhamer, J. J.
 TITLE Composition for the detection of signaling pathway gene expression
 JOURNAL Patent: US 6500938-A 1348 31-DEC-2002;
 FEATURES
 Location/Qualifiers
 1..3171
 source /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 3171
 Score: 829.00 Matches: 829
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-916-849A-1 (1-829) x AR270785 (1-3171)

QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuLeuLeuValCysTrpLeuGln 20
 DB 54 ATGGGGCTCCCTCGTGGAGCTCTCGGCTCTCCCTCTCTCCAGTTTGTGGGTGCAG 113
 QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
 DB 114 TGGCGGCTCCGAGCGGTGGCGGGGCTCTTCAGGAGGCTGAAGTACCTTGGAGGG 173
 QY 41 GlyValAlaGluGlnProGlyGlnAlaLeuGlyValPheMetGlyCysProGly 60
 DB 174 GGAGGCGGAGGAGGAGGCGCGGCGGCTGGGAAAGTATTATGGGCTGGCCCTGGG 233
 QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
 DB 234 CAAGACCCAGCTCTGTTAGCACTGATATGATGACTTCACTGTGCGGATGGCGAGACA 293
 QY 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
 DB 294 GTCCAGGAAGAAGGTCACTGAAGGAAGAAGAAATCCATTGAAGATCTTCCCATCCAAAGT 353
 QY 101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
 DB 354 ATCTTACGAAGACACAAAGAGATTGGGGTGTGTCTCCAAATCTGTCCCTGAAAATGGC 413
 QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
 DB 414 AAGGGTCCCTTCCCCAGAGACTGAATCAGCTCAAGTCTAATAAGATAGAGACACCAAG 473
 QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal 160
 DB 474 ATTTCTACAGATCATCGGGGCGGGGCGAGACACGCCCTTGGAGGGTGTCTTCGCTGTA 533
 QY 161 GluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
 DB 534 GAGAAGGACAGAGCTGGTGTGTTGTAATAGCCACTGGACCGGGAGGAGATTGCCAAG 593
 QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProValAsn 200
 DB 594 TATGAGCTCTTTGGCCACGCTGTGTGAGAAATGGTGTCTCAGTGGAGGACCCCATGAAC 653
 QY 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
 DB 654 ATCTCCATCATGTCGACCGACCAAGATGACCAAGACCCCAAGTTTACCCAGGACACCTTC 713

QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
 DB 714 CGAGGGAGTGTCTAGAGGGAGTCTTACCAGGTACTTCTGTGATGAGGTGAGACGCCACA 773
 QY 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
 DB 774 GATGAGGATGATGCATCTACACCTACATGGGGTGGTCTTACTTCACTCATAGCAA 833
 QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
 DB 834 GAACCAAGAGACCCACAGCCTCATTTCCAAATTCACCGGAGCAGACGACCATCAGC 893
 QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
 DB 894 GTCACTCTCCAGTGGCTGGACCGGAAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
 QY 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluIleLeuAsp 320
 DB 954 ACAGACATGGATGGGACGGCTCCACCACCGGAGTGGCAGTAGTGAGATCTTGTAT 1013
 QY 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
 DB 1014 GCCAATGACAAATGCTCCCAATGTTGACCCCAAGAGTACGAGGCCCATGTGCTGAGAT 1073
 QY 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
 DB 1074 GCAGTGGGCCATGAGGTGCAGAGGCTGACGGTCACTGATCTGGACGCCGCCCACTCACA 1133
 QY 361 AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
 DB 1134 GCGTGGGCTGCCACCTACCTTATCATGGGGGTGACGAGGGAGCCATTTTACCATCAC 1193
 QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
 DB 1194 ACCCACCTGAGAGCAACACAGGCGATCTTGACAAACACGAGGAGGTTTGGATTTGAGGCC 1253
 QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
 DB 1254 AAAAAACAGCACACCTGTACGTTGAAGTGACCAACAGGAGGCCCTTTGTGCTGAAGCTC 1313
 QY 421 ProThrThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
 DB 1314 CCAACCTCCACAGCCACCATAGTGTCCAGTGGAGGATGTGAATGAGCACCTGTGTTT 1373
 QY 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
 DB 1374 GTCCACCTCCAAAGTCGTTGAGGTCCAGGAGGCAATCCCCACTGGGAGGCTGTGTGT 1433
 QY 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuAsp 480
 DB 1434 GTCTACACTCCAGAAACCCCTGACAGGAGATCAAAAGATCAGCTACCGCATCTTGAGA 1493
 QY 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
 DB 1494 GACCCAGCAGGTGGCTAGCCATGGACCCAGACAGTGGGCGAGTCCAGCTGTGGGCACC 1553
 QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
 DB 1554 CTCGACCGTGAAGATGAGCAGTTGTGAGGAACAACATCTATGAAGTCATGTGCTTGGCC 1613
 QY 521 MetAspAsnGlySerProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeuAsp 540
 DB 1614 ATGGACAAATGGAAGCCCTCCACCACTGGCAGGGAACCTTCTGCTAAACACTGATTGAT 1673
 QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
 DB 1674 GTCAACGACCATGGCCAGTCCCTGAGCCCCGTCAGATCACCATCTGCAACCAAGCCCT 1733
 QY 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
 DB 1734 GTGGGCCGCTGTGAACATCAGGAGCAAGACCTGTCTCCCCACACCTCTCCCTTTCCAG 1793
 QY 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600


```

1794 GCCCAGTCCACAGATGACTCAGACATCTACTGACGCGCAGAGTCAACAGGAAGGTGAC 1853
QY ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
1854 ACATGGTCTGTCTCCCTGAGAGAGTTCCTGAGCAGAGATACATATGAGTGCACCTTTCT 1913
QY LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
1914 CTGTCTGACCATGTCACAAAGAGCAGCTGACGGTGTATCAGGCGCACTGTGTGCGACTGC 1973
QY HisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValLeu 660
1974 CATGCCATGTCGAAACCTGCCCTGACCTCGAAGAGGTTTCATCTCCCTGCTGCTG 2033
QY GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
2034 GGGCTGTCTCTGCTGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2093
QY ArgLysIleLysGluProLeuLeuLeuProGluAspThrArgAspAsnValPheTyr 700
2094 CGGAAGATCAAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2153
QY TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
2154 TATGCCAAGAGGGGGTGGCAAGAGGACCCAGACTATGACATCACCCAGCTCCACCGA 2213
QY GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
2214 GGTCTGAGGCCAGCCGCGAGGTGTTCTCCGAATGAGTGGCCACCAACCATCATCCCG 2273
QY ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleLeuGlu 760
2274 ACACCATGTACCGTCTAGCCGCCACCAACCCAGATGAAATCGCACTTTATATTTGAG 2333
QY AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
2334 AACCTGAGCGGGCTAACAGACCCACAGCCCGCCCTAGACACCTCTGTGTGTTC 2393
QY AspTyrGluGlySerGlySerAspAlaAsnLeuSerSerLeuThrSerSerAlaSer 800
2394 GACTATGAGCGAGCGCTCCGACGCGCTCCCTGAGCTCCCTACCTCCCTCCGCTCC 2453
QY AsnGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
2454 GACCAAGACCAAGATTACGATTATCTGAACAGTGGGGCAGCGCTTCAGGAAGCTGGCA 2513
QY AspMetTyrGlyGlyGlyGluAspAsp 829
2514 GACATGTACGGTGGCGGGAGGACGAC 2540

```

```

RESULT 2
AX411115 3171 bp DNA linear PAT 14-JUN-2002
LOCUS Sequence 3762 from Patent WO0229103.
DEFINITION AX411115
ACCESSION AX411115
VERSION AX411115.1 GI:21443820
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J. G.
Gene expression profiles in liver cancer
Patent: WO 0229103-A 3762 11-APR-2002;
JOURNAL GENE LOGIC INC (US)
FEATURES
Location/Qualifiers
1..3171
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="EMBL/GenBank Accession No. X63629"

```

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 3171
 Score: 829.00 Matches: 829
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-09-916-849A-1 (1-829) x AX411115 (1-3171)

QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln 20
 Db 54 ATGGGGCTCCCTCGTGACCTCTCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113
 QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
 Db 114 TGGCGGCTCCCGAGCGCTGGCGGGCTCTTCAGGAGGCTGAAGTGAACCTTGGAGGG 173
 QY 41 GlyGlyAlaGluGlnProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
 Db 174 GGAGGCGCGAGCAGGAGCCCGCCGCGCTGGGAAAAGTATTATGGGCTGCCCTGGG 233
 QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspAspPheThrValArgAsnGlyGluThr 80
 Db 234 CAAGAGCCACTCTGTTAGCACTGATATGATGACTTCTCTGTGCGGAATGGCGAGACA 293
 QY 81 ValGlnIleArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
 Db 294 GTCCAGAAAGAAAGGTCACTGAGGAAAGAAATCCATTGAAGATCTTCCCATCCAAAGCT 353
 QY 101 IleLeuArgArgHisLysArgAspTyrValValAlaProIleSerValProGluAsnGly 120
 Db 354 ATCTACGAGACACAAGAGAGATTGGGTGGTGTCTCTCAATATCTGCTCCTGAATGGC 413
 QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
 Db 414 AAGGGTCCCTTCCCGAGAGACTGAATCAGCTCAAGTCTTAATAAGATAGAGACCAAG 473
 QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal 160
 Db 474 ATTTCTACAGCATACGCGGGCGGGGCGAGACAGCCCTCCCTGAGGTGTCTTCCTGTA 533
 QY 161 GluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluIleAlaLys 180
 Db 534 GAGAAGGAGACAGCTGGTGTGTGTTGAATAAGCCACTGGACCGGGAGGAGATTGCCAAG 593
 QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValIleLysProMetAsn 200
 Db 594 TATGAGCTCTTTGGCCACGCTGTGTGAGAGATGTGCTCAGTGGAGGACCCCATGAC 653
 QY 201 IleSerIleLeuValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
 Db 654 ATCTCATCATCTGTGACCGAGAGTACCAAGCCCAAGTTTACCAGGACACCTTC 713
 QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
 Db 714 CGAGGAGTGTCTTAGAGGAGTCTTACCGAGTACTTCTGTGATGACGAGTGCACCCACA 773
 QY 241 AspGluAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
 Db 774 GATGAGGATGATGCCATCTACACCTACAAATGGGGTGGTGTCTTACTCCATCCATAGCAA 833
 QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
 Db 834 GAACCAAGGACCCACACCTCATGTTCAATTCACCGAGGACAGGACCACTACAGC 893
 QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
 Db 894 GTCATCTCCAGTGGCTGGACCGGAAAAAGTCCCTGTAGTACACACTGACCATCCAGGCC 953
 QY 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValIleGluLeuAsp 320

954	Db	ACAGACATGGATGGGGACGGCTCCACACACCGCAGTGGCAGTAGTGGAGATCCTTTGAT	1013
321	Qy	AlaAsnAspAsnAlaProMetPheAspProGlnIlystYrGluAlaHisValProGluAsn	340
1014	Db	GCCAAATGACAATGCTCCCATGTTTGACCCCCAGAAAGTACGAGGCCCATGTGCCTGAGAAT	1073
341	Qy	AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro	360
1074	Db	GCATGGGCCCATGAGGTGCAGAGCTGACGGTCACTATCTGGACGCCCACTACCA	1133
361	Qy	AlaTrpArgAlaThrTyrrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr	380
1134	Db	GCGTGGCGTGCCACCTACCTTATCATGGCGGTGACGACGGGACCAATTTACATCACC	1193
381	Qy	ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgIysGlyLeuAspPheGluAla	400
1194	Db	ACCACACCTTGAGACACACCGGCATCTGTACACACAGGAAGGTTTGGATTTTAGGCC	1253
401	Qy	LysAsnGlnHisThrLeuTyrrValGluValThrAsnGlnAlaProPheValIleuysLeu	420
1254	Db	AAAAACACGACACACCTTGAGTTGAAGTGACCAACGAGGCCCTTTTGTGTGAAGCTC	1313
421	Qy	ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe	440
1314	Db	CCAACTCCACAGCCACCATAGTGGTCCACGTGGAGGATGTGAATGAGGCACCTGTGTTT	1373
441	Qy	ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys	460
1374	Db	GTCCACCCCTCCAAAGTCGTGTAGGTCCAGAGGGCATGCCCACTGGGAGCCTGTGTGT	1433
461	Qy	ValTyrrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrrGileLeuAsq	480
1434	Db	GTCTCACTCCAGAGACCCCTGCACAGGAGATCAAAAGATCAGCTACCGCATCCTGAGA	1493
481	Qy	AspProAlaGlyTrpIleuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr	500
1494	Db	GACCCACAGAGGTGGCTAGCCATGTGACCCACAGACAGTGGCAGGTTCACAGTGTGGGCACC	1553
501	Qy	LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrrGluValMetValIleuAla	520
1554	Db	CTCGACCGTGAGATGAGCAGTTGTGTAGAGAACAAATCTATGAAGTCAATGGTCTTGGCC	1613
521	Qy	MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrIleuIleAsp	540
1614	Db	ATGAGCAATGTGAAGCCCTCCACCACTGGCAGCGGAAACCCCTTCTGTAAACACTGATTGAT	1673
541	Qy	ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro	560
1674	Db	GTCAACGACATGGCCCGATCCCTGTAGCCCGGTCAATCACCATCTGCACCAAGACCCCT	1733
561	Qy	ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln	580
1734	Db	GTGGCCACGTGCTGAACATCACGGACAAGGACCTGTCTCCCAACACCTCCCTTTCAG	1793
581	Qy	AlaGlnLeuThrAspAspSerAspIleTyrrTrpThrAlaGluValAsnGluGlyAsp	600
1794	Db	GCCAGCTCACAGATGACTCAGACATCTACTGGACGCGAGAGGTCAACGAGAAGGTGAC	1853
601	Qy	ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrrAspValHisLeuSer	620
1854	Db	ACAGTGTCTGTTCCTGAAGAAGTTCCTGAAGCAGGATACATATGACGTGCACCTTTCT	1913
621	Qy	LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys	640
1914	Db	CTGTCTGCACATGGCAACAAAGACAGAGCTGACCGGTATCAGGGGCCCATGTGTGTGCAC	1973
641	Qy	HisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValLeu	660
1974	Db	CATGGCCATGTGAACCTGCCCTGGACCTGGAAAGGAGGTTTTCATCTCCTCTGTGTG	2033
661	Qy	GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys	680

[illegible]

114 TCGCGGGCTCGAGCGCTGCCGGGCTTCAGGAGGCTGAAGTGAACCTTGAGGCG 173
41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyValPheMetGlyCysProGly 60
174 CGAGCGGGAGCAGAGCGCGCGCGAGCGCTGGGGAAGTATTCATGGCTGCGCTGG 233
61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
234 CAGAGACGAGCTCTGTAGCACTGATATGATGACTTCACTGTGCGGAATGGCGAGACA 293
81 ValGlnGluArgSerLeuLeuGluArgAsnProLeuLeuValPheProSerLeuArg 100
294 GTCCAGGAAGAGGTCACTGAAGGAAGGAATCCATTGAAGATCTCCCATCCAAAGCT 353
101 IleLeuArgArgHisLeuArgAspTrpValAlaAlaProIleSerValProGluAsnGly 120
354 ATCTTACGAGACACAGAGAGATGGGTGGTGTCTCCATATCTGTCTCTGAAATGGC 413
121 LysGlyProPheProGluArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
414 AAGGGTCCCTCCCGCAGAGCTGAATCAGCTCAAGTCTAATAAAGATAGACACCAAG 473
141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
474 ATTTTCTACAGCATCACGGGGCGGGGAGACAGCCCCCTGAGGGTCTTCTGCTGTA 533
161 GluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
534 GAGAGGAGACAGGCTGGTGTGTGTGAATAGCCACTGGACCGGAGGAGATTGCCAAG 593
181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
594 TATGAGCTTTGGCCACGCTGTGAGAGAAATGGTGTCTCAGTGGAGACCCCATGAAC 653
201 IleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
654 ATCTCATCATGTCGACCGACCAAGATGACCAAGCCCAAGTTTACCAGAGACCTTC 713
221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
714 CGAGGAGGTCTTAGAGGAGTCTTACCAGGTACTTCTGTGATGACGAGTGCACGCCACA 773
241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
774 GATGAGGATGATGCCATCTACACCTACAAATGGGGTGTTCCTTCTCATCCATAGCCAA 833
261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
834 GAACCAAGAGCCACACGACCTCATGTTCAAAATTCACCGAGACACGACCATCAGC 893
281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
894 GTCATCTCCAGTGGCTGGACCGGGAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValValGluIleLeuAsp 320
954 ACAGACATGGATGGGACGGCTCCACCAACCGGAGTGGCAGTAGTGGAGATCTTGAT 1013
321 AlaAspAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
1014 GCCAATGACAATGCTCCCATGTTTACCCCGGAGAGTACGAGGCCCATGTGCTGAGAAT 1073
341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
1074 GCAGTGGGCCATGAGGTGCAGAGCTGACGGTCACTGATCTGAGCGCCCCCAACTCACA 1133
361 AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
1134 GCGTGGCTGCCACCTTACCTTATCATGGCGGTGACGAGGGGACCATTTTACCATCACC 1193
381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400

1194 ACCACCTCGAGAGCAACACGAGGCTCTGACAAACAGGAGGGTTTGGATTTTGGGCC 1253
401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
1254 AAAAACCGACACACCTCTGTACGTTGAAGTGACCAACGAGGCCCTTTTGTGCTGAAGCTC 1313
421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
1314 CCAACCTCCACAGCACCATAGTGGTCCACGCTGGAGGATGTGAATGAGGCACCTGTGTT 1373
441 ValProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
1374 GTCCACCTCCCAAAGTCGTTGAGGTCAGAGGGCATCCCCACTGGGAGCCTGTGTGT 1433
461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
1434 GTCTACACTGCAGAGACCCCTGACAAAGAAATCAAAAGATCAGTACCACCTCTGAGA 1493
481 AspProIleGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
1494 GACCCAGCGGGTGGCTAGCCATGGACCCAGACAGTGGGAGGTACAGCTGGGGCACC 1553
501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
1554 CTCGACCTGAGGATGAGCAGTTTGTGAGGAACAACATCTATGAAGTCATGGTCTTGCC 1613
521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
1614 ATGACAAATGGAAGCCCTCCACACTGGCAGCGGAACCCCTCTGTCTAACTGATGAT 1673
541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
1674 GTCACCAACCATGGCCCGGCTCCCTGAGCCCGCTCAGATCACCATCTGCACCAAGGCCCT 1733
561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
1734 GTGGCCACCTGTGTAACATCAGGACAAGGACCTGTCTCCCAACACCTCCCTTTCAG 1793
581 AlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGluAsp 600
1794 GCCAGCTCACAGATGACTCAGACATCTACTGACCGCAGAGGTCAACGAGAGGTGAC 1853
601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
1854 ACAGTGTCTTGTCCCTGAAGAGTTCTCTGAAGCAGGATACATATGACGTGCACCTTCT 1913
621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
1914 CTGTCTGACCATGGCAACAAGAGACAGCTCAGCGTGTATCAGGGCCACTGTGTGGACTGC 1973
641 HisGlyHisValGluThrCysProGlyProTriLysGlyGlyPheIleLeuProValLeu 660
1974 CATGCCCATGTCGAAACCTGCCCCCTGGACCTGGAAAGGAGGTTTCATCCTCCCTGTGCTG 2033
661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
2034 GGGGCTGTCTGGCTGTGCTGTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2093
681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
2094 CGGAAGATCAAGAGACCCCTCTACTCCGAGAAGATGACCCCGTGAACAGCTTCTTCTAC 2153
701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
2154 TATGGCAAGAGGGGGTGGCGAAGAGGACCAAGGACTATGACATCACCAGCTCCACCGA 2213
721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
2214 GGTCTGAGGCGCAGCGCGAGGTGTCTCCGCAATACGCTGGGACCAACCATCATCCCG 2273
741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleGlu 760
2274 ACACCATGTACCTCTCTAGCCAGCCAAACCCAGATGAATCGGCACTTTATAATTGAG 2333

```

Qy 761 AsnLeuIysAlaIasThrAspProThrAlaProProTyAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGCGCGCTAACACAGACCCACAGCGCCCGCTTAGCACCCCTCTGTGTGTC 2393
Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGCGAGCGGCTCCGAGCGCGGTCCCTGAGCTCCCTACCTCCCTCGCGCTCC 2453
Qy 801 AspGluAspGlnAspTyrAspTyrLeuAsnGluTrrpGlySerArgPheIysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTCAACAGAGTGGGCGAGCGGCTTCAAGAAGCTGGCA 2513
Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 2514 GACATGTACGTCGCGGGGAGGACGAC 2540

RESULT 4
LOCUS HSPCAD 3171 bp mRNA linear PRI 06-FEB-1992
DEFINITION H sapiens mRNA for p cadherin.
ACCESSION X63629
VERSION X63629.1 GI:35322
KEYWORDS cadherin.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3171)
AUTHORS Shimoyama,Y., Yoshida,T., Terada,M., Shimosato,Y., Abe,O. and Hirohashi,S.
TITLE Molecule cloning of a human Ca2+-dependent cell-cell adhesion molecule homologous to mouse placental cadherin: its low expression in human placental tissues
J. Cell Biol. 109 (4 Pt 1), 1787-1794 (1989)
JOURNAL 90009051
MEDLINE 2793940
FEATURES
     source          Location/Qualifiers
     1..3171
     /organism="Homo sapiens"
     /mol_type="mRNA"
     /db_xref="taxon:9606"
     54..2543
     /codon_start=1
     /product="p-cadherin"
     /protein_id="CAA45177.1"
     /db_xref="GI:35323"
     /db_xref="GOA:P22223"
     /translation="MGLPRGLASILLQVLCQAASEPCRAVFRAEVILEAGAE
     QEPQALGVKFCQPGQEPALFSTNDQFTVRNGSTQVERSLKERNPKLIPPSKRIL
     RHKQDWVAPISVPENGKPPORLNQKSNKRDTKIFYITPGADSPPEGVFV
     EKETGWLKLPDLREETAKYELFGHAYSENGASVEDPMNISIIIVTDQNDHKPFTQD
     TFRGSVLGVLPGTSMQVATDDEDAIYTVNGVAYSIIHSEOPKPDHLMFTIHRST
     GTIIVSSGLDRKVEYVLTIIQATMDGDGSTTTAVAVELLDANDNAPWDFPKYE
     AVPENAVGHVQRVLTLDLAPNSPAMRATYLLINGDDGDHFTITTHPESNQGLIT
     RGLDFEAKQHILIVETNEAPVFKLPTSTATIVHVEDVNEAPVFPVSKVEVQ
     EGIPTEPCVYTAEDPKENOKISYRILRDPAGLWMDPDSQQVATVGTLDREDEQF
     VARNIYEVMLVANDGSPPTGTGTLTLLIDVNDHGPVPPEPQITICNQSPVRHLN
     ITDKOLSPITSFQALTDSDSIYWTAVNEBEGDTVLSLKFLKQDTYDVHLSLKH
     GNKEQLTVRAVTCCHGVETCPGPWKGGLTLPVLGVALLLFLLVLLLVKRRK
     IPEPLLPRDDRDNVFYVYEGSGGEEQDYDITOLHREGLARPEWLRNVDVPIIP
     TWYSPRPANPEIGNFIENLKNATPTAPTAPFTLLVFDYEGSGSDAASLSLTS
     ASDQDDYDLNWSGRFKLADNYGGGDD"
     3162..3167
     polyA_signal
     ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 3171
Score: 829.00 Matches: 829
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0

```

```

DB: 9 Gaps: 0
US-09-916-849A-1 (1-829) x HSPCAD (1-3171)
Qy 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCystTrpLeuGln 20
Db 54 ATGGGGCTCCCTCGTGACCTCTCGCGTCTCTCTCTCTCTCCAGGTTTCTGGCTGCAG 113
Qy 21 CysAlaIasSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
Db 114 TGC CGCGCTCCGAGCGCTCGCGCGGTCTTCAGGAGGCTGAAGTGACCTTGGAGGG 173
Qy 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db 174 GGAGCGCGAGGAGGAGCGCGCGCGCTGGGANAAGTATTTCATGGCTGCTGGG 233
Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
Db 234 CAAGAGCGAGCTGTGTTAGCACTGATAATGATCACTTCACTGTGGGAATGCGAGACA 293
Qy 81 ValGlnGluArgGserLeuLysGluAtqAsnProLeuLysIlePheProSerLysArg 100
Db 294 GTCCAGGAAGAAGGTCACTGAAGGAAGGAATCCATTGAAGATCTCCCATCCAAAGCT 353
Qy 101 IleLeuArgArgHisLysArgAspTrrpValValAlaProIleSerValProGluAsnGly 120
Db 354 ATCTTACGAAGACACAAAGAGAGATTGGGTGGTTCCTCAATATCTGCTCCCTGAAATGGC 413
Qy 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
Db 414 AAGGGTCCCTTCCCGCAGAGACTGAATCAGCTCAAGTCTAATAAAGATAGAGACCAAG 473
Qy 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal 160
Db 474 ATTTTCTACAGCATCAGCGGGCGGGGCGAGACAGCCCCCTGAGGGTGTCTTCGTGTA 533
Qy 161 GluLysGluThrGlyTrpLeuLeuLeuLysPheLeuAspArgGlnGluIleAlaLys 180
Db 534 GAGAAGGAGACAGCGCTGTGTTGTAATAAGCCACTGACCGGAGGAGATTGCGAAG 593
Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 594 TATGAGCTCTTTGGCCACGCTGTGTGAGAAATGGTCCCTCAGTGAGGAGCCCCATGAC 653
Qy 201 IleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 654 ATCTCCATCATCTGACCGCAGACCAAGATGACCAAGCCCAAGTTTACCCAGGACACCTTC 713
Qy 221 ArgGlySerValLeuGluValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 714 CGAGGGAGTGTCTTAGAGGAGTCTTACAGAGTACTTCTGTGATGCGAGTGACAGCCACA 773
Qy 241 AspGluAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
Db 774 GATGAGGATGATGCCATCTACACTCAATGGGTGGTGTCTTACTCATCCATGACCA 833
Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 834 GACCAAGAGACCCACACAGACCTCATGTTTCAATTCACCGGAGCAGAGCACCATCAGC 893
Qy 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db 894 GTCATCTCCAGTGGCTGGACCGGGAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
Qy 301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValValGluIleLeuAsp 320
Db 954 ACAGACATGGATGGAGCGGCTCCACCAACCGGAGTGGAGTAGTGAGATCTTGAT 1013
Qy 321 AlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1014 GCCAATGACATGCTCCCATGTTGACCCCGAGAGTACAGGCCCATCTGCTGTAGAT 1073
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360

```


Db 270 CAAAGCCAGCTCTGTTTAGCAGCTGATATGATGACTTCACTGTGCGGATGGCAGACA 329
 Qy 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
 Db 330 GTCCACGAAGAAGGTCACCTGAAGGAAGAAATCCATTGAAGATCTTCCCATCAAAAGT 389
 Qy 101 IleLeuArgHisLysArgAspThrValAlaProIleSerValProLusnGly 120
 Db 390 ATCTTACGAAGACACAAAGAGATGGGTGGTGTCTCCAAATATCTGTCCCTGAAAATGGC 449
 Qy 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
 Db 450 AAGGTCCTTCCCCAGAGACTGAATCAGCTCAAGTCTAATAAGATAGACACCAAG 509
 Qy 141 IlePheTySerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal 160
 Db 510 ATTTTCTACAGCATCACGGGGCGGGGAGACAGACCCCTGAGGGTGTCTTCCTGTA 569
 Qy 161 GluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
 Db 570 GAGAGAGAGACAGGCTGGTGTGTTGAATAGCCACTGACCGGGAGAGATGCCAAG 629
 Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
 Db 630 TATGAGCTCTTGGCCACGCTGTGTGAGAGATGGTGCCTCAGTGGAGAGACCCCATGAAC 689
 Qy 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
 Db 690 ATCTCCATCATCGTACCGACAGATGACCAAGCCCAAGTTTACCCAGGACACCTTC 749
 Qy 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
 Db 750 CGAGGGAGTGTCTTACAGGGAGTCTTACCAGGTACTTCTGTGTGATGAGGTGACGCCACA 809
 Qy 241 AspGluAspAspAlaIleTyThrThrAsnGlyValValAlaIleTySerIleHisSerGln 260
 Db 810 GATGAGGATGATCCCATCTACACCTCAATGGGGTGGTGTCTTACTCCATCCATAGCCAA 869
 Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
 Db 870 GAACCAAGAGACCCACACGACCTCATGTTCAATTCACCGGAGCACAGGCACCATCAGC 929
 Qy 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyThrIleThrIleGlnAla 300
 Db 930 GTCATCTCCAGTGGCCTGGACCGGGAAAAAGTCCCTGAGTACACACTGACCATCCAGGCC 989
 Qy 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValAlaValGluIleLeuAsp 320
 Db 990 ACAGATGATGGGACCGCTCCACACACGAGCTGAGTGGAGTGGAGATCTTGTAT 1049
 Qy 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyGluAlaHisValProGluAsn 340
 Db 1050 GCCAATGACAATCTCCATGTTTACCCGCCAGAGTACGAGGCCCATGTGCTGAGAT 1109
 Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
 Db 1110 GCAGTGGGCGCATGAGGTGACGAGCTGAGGTCTGATCTGACGCCCCCACTACCA 1169
 Qy 361 AlaTrpArgAlaThrTyThrIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
 Db 1170 GCGTGGGTCGCCACCTTATCATGGGCGGTGACGAGCGGACCATTTTACCATCACC 1229
 Qy 381 ThrHisProGluSerAsnGlnIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
 Db 1230 ACCCACCTGAGAGCAACCCAGGCACTCTGACAAACCCAGGAAGGTTTGGATTGAGGCC 1289
 Qy 401 LysAsnGlnHisThrLeuTyValGluValThrAsnGluAlaProPheValLeuLysLeu 420
 Db 1290 AAAAAACGACACACCTGTGTGAGTGTGAGTGACCAACGAGGCCCTTTTGTGTGAGGCTC 1349
 Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440

Db 1350 CCAACCTCCACAGCCACCACCATAGTGGTCCAGCTGAGGATGTGAATGAGGCACCTGTGTTT 1409
 Qy 441 ValProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
 Db 1410 GTCCACCTCCAAAGTCGTGAGGTCCAGGAGGCATCCCCACTGGGAGCCTGTGTGT 1469
 Qy 461 ValTyThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyArgIleLeuArg 480
 Db 1470 GTCTACACTGACAGAGACCTGACAGGAGAATCAAAAGATCAGCTACCGCATCTGAGA 1529
 Qy 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
 Db 1530 GACCCAGCAGGGTGGCTAGCCATGGACCCACAGAGTGGGAGGTCAAGCTGTGGGCACC 1589
 Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyThrGluValMetValLeuAla 520
 Db 1590 CTGACCGTGAAGATGAGCAGTTTGTGAGGAACAATCATCTATGAAGTCATGGCTTGGCC 1649
 Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuLeuLeuLeuLeu 540
 Db 1650 ATGGACAATGGAAGCCCTCCACCACTGGCAGCGGAACCCCTCTGTCTAAGTGTAT 1709
 Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
 Db 1710 GTCAACGACCATGGCCAGTCCCTGAGCCCGCTCAGATCAACATCTGCAACCAAGCCCT 1769
 Qy 561 ValArgHis-Val-LeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheG 580
 Db 1770 GTGGCCCA-GTGGCTGAACATCACGACAAGGACCTGTCTCCCAACACCTCCCTTCC 1828
 Qy 580 InAlaGlnLeuThrAspAspSerAspIleTyThrThrAlaGluValAsnGluGluGlyA 600
 Db 1829 AGGCCACAGCTCACAGATGACTCAGACATCTACTGGACGGCAGAGGTCAACAGAGAAGTG 1888
 Qy 600 spThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyThrValHisLeuS 620
 Db 1889 ACACAGTGGTCTTGTCCCTGAAGAGTCTCTGAAGCAGGATACATATGACGTGCACCTTT 1948
 Qy 620 erLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspC 640
 Db 1949 CTCTGTCTGACCTGGCAACAAGAGCAGTGCAGGTGATCAGGCGCATCTGTGTGGACT 2008
 Qy 640 YHisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValL 660
 Db 2009 GCCATGGCCATGTCGAACCTGCTGGACCTTGGAGGGAGGTTTCATCTCTCTGTGC 2068
 Qy 660 euGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysL 680
 Db 2069 TGGGGGTGTCTCTGCT 2128
 Qy 680 YsArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPhe 700
 Db 2129 AGCGAAGATCAAGAGGCCCT 2188
 Qy 700 YrTyGlyGluGluGlyGlyGlyGluAspGlnAspTyThrAspIleThrGlnLeuHis 720
 Db 2189 ACTATGCGAAGAGGGGGTGGCGAGAGGACAGGACTATGACATCACCCAGCTCCACC 2248
 Qy 720 rgGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIleP 740
 Db 2249 GAGTCTGTGAGGGCCAGGCCCGGAGTGGTCTCTCGCAATGACGTGGCAGCAACCATCATCC 2308
 Qy 740 roThrProMetTyArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleG 760
 Db 2309 CGACACCATGATCCGCT 2368
 Qy 760 luAsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyThrAspThrLeuValP 780
 Db 2369 AGAACCTGAAGGGCGCTTAACACAGACCCCAAGCCCGCCCTACGACACCTCTTGGTGT 2428
 Qy 780 heAspTyThrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAla 800
 Db 2429 TCGACTATGAGGGCAGGGGCTCGACCGCCGCTCCCTGAGCTCCCTCACCTCTCCGCTC 2488

QY 81 ValGlnGluArgSerLeuLysGluAsgAsnProLeuLysIlePheProSerLysArg 100
DB 785 GTCCAGGAAGAAGGTCACTGAGGAAAGGAATCCATTGAGATCTTCCCATCCAAACGT 844
QY 101 IleLeuArgArgHisLysArgAspTrpValAlaProIleSerValProGluAsnGly 120
DB 845 ATCTTACGAAGACACAAAGAGAGATGGGTGGTTCCTCCCAATATCTGCTCCCTGAAATGGC 904
QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
DB 905 AAGGGTCCCTTCCCCAGAGACTGAATCAGCTCAGTCTAATAAGATAGACACCAAG 964
QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal 160
DB 965 ATTTTCTACAGCATCACGGGCGCGGGCGAGACAGCCCCCTGAGGGTGTCTTCGCTGTA 1024
QY 161 GluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
DB 1025 GAGAAAGAGACAGGCTGGTGTGTTGTAATAAGCACTGGACCGGAGAGATTCGCAAG 1084
QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
DB 1085 TATGAGCTCTTGGCCACCGCTGTGTGAGAAATGGTGCCTCAGTGGAGGACCCCATGAAC 1144
QY 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
DB 1145 ATCTCCATCATAGTACCGACCAAGATGACCAAGCCCAAGTTTACCCAGGACACCTTC 1204
QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnVal-ThrAlaLys 240
DB 1205 CGAGGGAGTGTCTTAGGGAGGTCCTACCGAGTACTTCTGTGATGCAGAT-GACAGGCCAC 1263
QY 240 AspGluAspAspAlaIleTyrThrTyrAsnGlyValAlaIleTyrSerIleHisSerGln 260
DB 1264 AGATGAGGATGATCCATCTACACCTACCAATGGGTGGTGTCTTACTCCATCCATAGCCA 1323
QY 260 nGluProLysAspProHisLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
DB 1324 AGAACCAAGAGACCCACACACCTCATGTTCACATTCACCGAGACAGGACCATCAG 1383
QY 280 rValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAl 300
DB 1384 CGTCATCTCCAGTGGCTGGACCGGGAAGTCCCTGAGTACACACTCACCATCCAGGC 1443
QY 300 aThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValGluIleLys 320
DB 1444 CACAGACATGGATGGGAGCGGCTCCACCCAGAGTACGAGGCCCATGTGCTGAGAA 1503
QY 320 pAlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAs 340
DB 1504 TGCCATGACAAATGCTCCCATGTTTGACCCCAAGATACGAGGCCCATGTGCTGAGAA 1563
QY 340 nAlaValGlyHisGluValGlnArgIleThrValThrAspLeuAspAlaProAsnSerPr 360
DB 1564 TGCAGTGGGCGCATGAGTGCAGAGGTGACGGTCACTGATCTGGACGCCCACTCAC 1623
QY 360 oAlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleTh 380
DB 1624 AGCGTGGCGTGCACCTACCTATCATTTGGCGGTGACGAGGGGACCATTTTACCATCAC 1683
QY 380 rThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAl 400
DB 1684 CACCCACCTTGAGAGCAACCCAGGCACTCTGACAAACAGGAAGGGTTTGGATTTTGAGGC 1743
QY 400 aLysAsnGlnHisThrIleTyrValGluValThrAsnGluAlaProPheValLeuLysIle 420
DB 1744 CAAAAACCAAGACACCTGTACCTTGAAGTGAACCAACAGGGCCCCCTTTTGTGCTGAGCT 1803
QY 420 uproThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPh 440
DB 1804 CCCAACCTCCACAGCCACCATAGTGTGTCACGTGGAGGATGTAATGAGGACCTGTGT 1863

QY 440 eValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCy 460
DB 1864 TGTCCCACTCCCAAGTCTGTAGGTTCAGGAGGGCATCCCACTGGGAGGCTGTGTG 1923
QY 460 sValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuAr 480
DB 1924 TGTCTACACTGCAGAGACCTTCACAGGAGATCAAAAGATCAGCTACCGCATCTGAG 1983
QY 480 gAspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyTh 500
DB 1984 AGACCCAGCAGGCTGGCTAGCCATGGACCCAGACAGTGGGCAGGTACAGCTGTGGGCAC 2043
QY 500 rLeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAl 520
DB 2044 CCTCGACCGTGAGGATGAGCAGTTTGTGAGGAAACAATCTATGAAATCATGTGCTTGGC 2103
QY 520 aMetAspAsnGlySerProProThrThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAs 540
DB 2104 CATGACAAATGGAAAGCCCTCCACCACTGGACGGGACCCCTTCTGCTAACACTGATTGA 2163
QY 540 pValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPr 560
DB 2164 TGTCAACGACCATGGCCAGTCTCTGAGCCCGCTCAGATCACCATCTGCAACCAAGCCC 2223
QY 560 oValArgHis-ValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheG 580
DB 2224 TGTGGGCCA-GGTGCTGAACATCAGGACAAGACCTGTCTCCCAACCTCCCTTTC 2282
QY 580 lAlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyA 600
DB 2283 AGGCCAGCTCACAGATGACTCAGACATCTACTGGACGGCAGAGGTCAACAGGAAAGTG 2342
QY 600 spThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHisLeuS 620
DB 2343 ACACAGTGGTCTGTCCCTGAAGAAGTTCTGAAGCAGGATACATATGAGTGCACCTTT 2402
QY 620 erLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspC 640
DB 2403 CTCTGTCTGACCATGGCAACAAGAGCAGCTGACGGTGAATCAGGGCCACTGTGTGGACT 2462
QY 640 ySHisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProVal 660
DB 2463 GCCATGCCATGTGCAAACTGCGCTGGACCTGGAAAGAGGTTTTCATCCTCCCTGTGC 2522
QY 660 euGlyAlaValLeuAlaLeuPheLeuLeuValLeuLeuLeuValArgLysL 680
DB 2523 TGGGGGTGTCTTGGCTCTGTCTCTCTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2582
QY 680 ySArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheT 700
DB 2583 AGCGGAAGATCAAGAGGCCCTCTCTACTCCAGAAAGATGACACCCGCTGACACGCTTCT 2642
QY 700 yTrTyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisA 720
DB 2643 ACTATGGCGAAGAGGGGGTGGCGAAGAGGACCAAGACTATGACATCACCCAGCTCCACC 2702
QY 720 rGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIleP 740
DB 2703 GAGTCTGGAGGCCAGGCCGGGTGGTCTCTCCGAATGACGTGGGACCAACCATCATCC 2762
QY 740 roThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleG 760
DB 2763 CGACACCATGATACCGTCTTAGGCCAGCAACCCAGATGAATCGGCACTTTATTAATTG 2822
QY 760 luAsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValP 780
DB 2823 AGAACCTGAAGGGCGCTAACACAGACCCCAAGCCCGCCCTACGACACCCCTCTTGGTGT 2882
QY 780 heAspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlas 800
DB 2883 TCAGCTATGAGGGCAGCGGCTCCGACCGCGCTCCCTGAGCTCCCTCACCTCTCCGCT 2942
QY 800 erAspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLeuA 820

Db 2943 CCGACCAACACAGATACAGTATCTGACGAGTGGGCGAGCGCTTCAAGAGCTGG 3002
QY 820 laAspMetTyrglyGlyGluAsp 829
Db 3003 CAGACATGACGTTGGCGGGGAGGACGAC 3031
RESULT 7
AX210689 306 bp DNA linear PAT 31-AUG-2001
LOCUS
DEFINITION Sequence 331 from Patent WO0157058.
ACCESSION AX210689
VERSION AX210689.1 GI:15424949
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 Rosenthal, A., Hinzmann, B., Schaefer, R., Zuber, J., Tchernitsa, O.,
Grips, M., Hellriegel, M., Schmitz, A.C. and Sers, C.
TITLE Detection of differential gene expression
JOURNAL Patent: WO 0157058-A 331 03-AUG-2001;
Metagen Gesellschaft fuer Genomforschung mbH (DE)
FEATURES
source
1. 306
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 3.46e-98 Length: 306
Score: 102.00 Matches: 102
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 12.30% Indels: 0
DB: 6 Gaps: 0
US-09-916-849A-1 (1-829) x AX210689 (1-306)
QY 707 GlyGluGluAspGlnAspTyraAspThrGlnLeuHisArgGlyLeuGluAlaArgPro 726
Db 1 GCGAGAGAGCAGCAGGACTATGACATCCACCCAGCTCCACCGAGGCTGGAGGCCAGCGCG 60
QY 727 GluValValLeuArgAsnAspValAlaProThrilleProThrProMetTyraArgPro 746
Db 61 GAGGTGGTTCTCCGCAATCAGCTGGCACCACCATCATCCCGACACCCCATGTACCGTCT 120
QY 747 ArgProAlaAsnProAspGluileGlyAsnPhellelleGluAsnLeuIysAlaAlaAsn 766
Db 121 CGGCCAGCCCAACCCAGATGAATCGCAACTTATTAATTGAGAACCTGAGCGGCTTAAC 180
QY 767 ThrAspProThrAlaProProTyraAspThrLeuValPheAspTyrglyGlySerGly 786
Db 181 ACAGACCCACAGCCCGCCCTACGACACCCCTCTTGGTGTCTGACTATGAGGCGAGCGGC 240
QY 787 SerAspAlaSerLeuSerSerLeuThrSerSerAlaSerAspGlnAspGlnAspTyra 806
Db 241 TCCGACGCGCGCTCCCTGAGCTCCCTCACCTCTCCGCTCCGACCAACCAAGATTAC 300
QY 807 AspTyr 808
Db 301 GATTAT 306
RESULT 8
AC099314 184945 bp DNA linear PRI 05-APR-2003
LOCUS
DEFINITION Homo sapiens chromosome 16 clone RP11-354W1, complete sequence.
ACCESSION AC099314
VERSION AC099314.3 GI:29568028
KEYWORDS
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 184945)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
REFERENCE
AUTHORS 2 (bases 1 to 184945)
DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE
AUTHORS 3 (bases 1 to 184945)
DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (04-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE
AUTHORS 4 (bases 1 to 184945)
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Submitted (05-APR-2003) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Apr 5, 2003 this sequence version replaced gi:18057085.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center and Los Alamos
National Laboratory
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.4.
FEATURES
source
1. 184945
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-354W1"
ORIGIN
Alignment Scores:
Pred. No.: 3.07e-72 Length: 184945
Score: 80.00 Matches: 80
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.65% Indels: 0
DB: 9 Gaps: 0
US-09-916-849A-1 (1-829) x AC099314 (1-184945)
QY 395 GlyLeuAspPheGluAlaLysAsnGlnHisThrLeuTyraValGluValThrAsnGluAla 414
Db 28107 GGTTTGGATTTTGAGGCCAAACACACACACCTGTACCTGTACCTGACCTACGAGGCC 28166
QY 415 ProPheValLeuLysLeuProThrSerThrAlaThrIleValHisValGluAspVal 434
Db 28167 CCTTTTGTGCTGAGCTCCCAACCTCCACAGCCACCATAGTGTCCACGTGGAGGATGTG 28226
QY 435 AsnGluAlaProValPheValProSerTyraValValGluValGlnGluGlyIlePro 454
Db 28227 AATGAGGCACCTGTGTTGTGTCCACCTCCAAAGTCCTTGGGTCCAGGAGGATCCCC 28286
QY 455 ThrGlyGluProValCysValTyThrAlaGluAspProAspTyraGluAsnGlnLysIle 474
Db 28287 ACTGGGAGGCTGTGTGTGTCTACCTGCAGAGACCTCCACAGGAGAGATCAAGATC 28346
RESULT 9
CNS01DXD/c
ID CNS01DXD standard; genomic DNA; HTG; 186939 BP.
XX
AC AL139227;
XX

SV AL139227.2
 XX 11-FEB-2000 (Rel. 62, Created)
 DT 25-MAY-2000 (Rel. 63, Last updated, Version 2)
 XX
 DE Human chromosome 16 DNA sequence *** IN PROGRESS *** BAC R-61512 of library
 DE RPCI-11 from chromosome 16 of Homo sapiens (Human)
 XX
 KW HTGS_DRAFT; HTGS_PHASE2.
 XX
 XX Homo sapiens (human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 XX
 RN [1]
 RP 1-186839
 RA Genoscope;
 RT Submitted (23-MAY-2000) to the EMBL/GenBank/DDBJ databases.
 RL
 XX

IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage, etc. . even if efforts are made to eliminate these contaminating sequences. The following BAC sequence is oriented from the T7 to the SP6 end.
 CC Contigs composition :
 CC 49057 bp contig from 1 to 49057
 CC 31448 bp contig from 49158 to 80605
 CC 55190 bp contig from 80706 to 135895
 CC 3703 bp contig from 135996 to 139698
 CC 5159 bp contig from 139799 to 144957
 CC 3046 bp contig from 145058 to 148103
 CC 14904 bp contig from 148204 to 163107
 CC 23632 bp contig from 163208 to 186839
 XX

 Overall quality chart :
 CC Range : bases
 CC 0 : 1083
 CC 1 - 9 : 1569
 CC 10 - 19 : 2752
 CC 20 - 29 : 6935
 CC 30 - 39 : 18862
 CC 40 - 49 : 16913
 CC 50 - 59 : 20166
 CC 60 - 69 : 35143
 CC 70 - 79 : 45986
 CC 80 - 89 : 26801
 CC 90 - 99 : 10629
 CC
 CC Percentage of bases with a quality value >= 40 : 83 %

 Key Location/Qualifiers
 FH source
 FT 1. 186839
 FT /chromosome="16"
 FT /db_xref="taxon:9606"
 FT /mol_type="genomic DNA"
 FT /organism="Homo sapiens"
 FT /clone_lib="RPCI-11"
 FT /clone="R-61512"
 XX
 SQ Sequence 186839 BP; 49577 A; 44657 C; 43147 G; 48728 T; 730 other;

Alignment Scores:
 Pred. No.: 3,1e-72 Length: 186839
 Score: 80.00 Matches: 80
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 9,65% Indels: 0
 DB: 30 Gaps: 0

US-09-916-849A-1 (1-829) x CNS01DXD (1-186839)
 QY 395 GlyLeuAspPheGluAlaLysAsnGlnHisThrLeuTyrValGluValThrAsnGluAla 414
 DB 18582 GGTTCGATTTCAGGCCAAACACACACACCTGTACCTGTGAAGTGACCAACGAGGCC 18523
 QY 415 ProPheValLeuLysLeuProThrSerThrAlaThrIleValHisValGluAspVal 434
 DB 18522 CCTTTGTGCTGAAGCTCCCAACCTCCACAGCCACCATAGTGTGCACGTGGAGGATGTG 18463
 QY 435 AsnGluAlaProValPheValProProSerLysValValGluValGlnGluGlyIlePro 454
 DB 18462 AATGAGGCACCTGTGTTGTTCCTCCACCTCCAAAGTGTTCAGGTCCAGGAGGCATCCCC 18403
 QY 455 ThrGlyGluProValCysValTyrThrAlaGluAspProAspLysGluAsnGlnLysIle 474
 DB 18402 ACTGGGAGCCTGTGTGTGTCTACTGTCAGAGACCTGCAAGGAGATCAAAAGATC 18343
 RESULT 10
 AC126773 196911 bp DNA linear PRI 26-NOV-2003
 LOCUS Homo sapiens chromosome 16 clone RP11-61512, complete sequence.
 DEFINITION AC126773 AL139227
 ACCESSION AC126773.3 GI:38524628
 VERSION AC126773.3
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 196911)
 AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 196911)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (09-JUL-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 3 (bases 1 to 196911)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (23-DEC-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 4 (bases 1 to 196911)
 AUTHORS DOE Joint Genome Institute
 TITLE Direct Submission
 JOURNAL Submitted (26-NOV-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT On or before Nov 26, 2003 this sequence version replaced GI:9212225, gi:27363207.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center and Los Alamos National Laboratory
 www.sngc.stanford.edu
 Quality: Phrap Quality >=40 100% of Sequence;
 Estimated Total Number of Errors is 0.
 Location/Qualifiers
 1. 196911
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-61512"
 FEATURES
 source
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,25e-72 Length: 196911
 Score: 80.00 Matches: 80
 Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00%
Query Match: 9.65%
DB: 9
Indels: 0
Gaps: 0

US-09-916-849A-1 (1-829) x AC126773 (1-196911)

QY 395 GlyLeuAspPheGluAlaLysAsnGlnHisThrLeuYrValGluValThrAsnGluAla 414
DB 178220 GATTTGGATTGAGGCCAAACACAGCACACCTGTACGTTGAAGTACCAACAGAGGCC 178279
QY 415 ProPheValLeuLysLeuProThrSerThrAlaThrIleValValHisValGluAspVal 434
DB 178280 CCTTTTGTCTGAAGCTCCCAACCTCCACAGCACCTAGTGTCCACGTCGAGGATGTG 178339
QY 435 AsnGluAlaProValPheValProProSerLysValValGluValGlnGluGlyIlePro 454
DB 178340 AATGAGGACCTGTGTTGTCCACCCCTCCAAAGTCGTTGAGTCCAGAGGGCATCCCC 178399
QY 455 ThrGlyGluProValCysValYrThrAlaGluAspProAspLysGluAsnGlnLysIle 474
DB 178400 ACTGGGGAGCCTGTGTGTCTACACTGCAGAGAGACCTGTGACAGGAGATCAAAAGATC 178459

RESULT 11
AP002024
LOCUS 54207 bp DNA linear HTG 01-JUN-2000
DEFINITION Homo sapiens chromosome 16 clone 97G12 map 16q22, *** SEQUENCING IN
PROGRESS ***, 3 ordered pieces.
ACCESSION AP002024
VERSION HTG; HTGS_PHASE2
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Tsai, S.F.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2000) Shih-Feng Tsai, National Yang-Ming
University, Institute of Genetics, 155 Li-Rong St. Section 2,
Pitou, Taipei, Taiwan 11221, Republic of China
(E-mail:lympet@ym.edu.tw, URL:http://genome.ym.edu.tw/,
Tel:886-2-28267043, Fax:886-2-28264930)

COMMENTS
* NOTE: This is a 'working draft', sequence as 100 N.
* consists of 3 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 1 15198: contig of 15198 bp in length
* 15199 15298: gap of 100 bp
* 15299 25043: contig of 9745 bp in length
* 25044 25143: gap of 100 bp
* 25144 54207: contig of 29064 bp in length.

FEATURES
Location/Qualifiers

1..54207
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q22"
/clone="97G12"

ORIGIN

Alignment Scores:
Pred. No.: 3,09e-64 Length: 54207
Score: 72.00 Matches: 72
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 8.69% Indels: 0

DB: 2 Gaps: 0
US-09-916-849A-1 (1-829) x AP002024 (1-54207)

QY 524 GlySerProProThrThrGlyThrGlyThrLeuLeuThrLeuLeuAspValAsnAsp 543
DB 292 GGAAGCCCTCCACCTGACGACGGAACCCCTTCTGTACACTGATTGATGACACGAC 351
QY 544 HisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerProValArgHis 563
DB 352 CATGGCCCACTCCTCGAGCCCGCTCAGATCACCATCTGCAACCAAGCCCTGTGGCCAC 411
QY 564 ValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGlnAlaGlnLeu 593
DB 412 GTGTGAACATCAGGCAAGGACCTGTCTCCACACACCTCCCTTCTCCAGGCCAGCTC 471
QY 584 ThrAspAspSerAspIleTyrThrAlaGluVal 595
DB 472 ACAGATGACTCAGACATCTACTGACGGCAGAGGTC 507

RESULT 12

BTCDHP
LOCUS 2106 bp mRNA linear MAM 20-JUL-1995
DEFINITION Bovine mRNA for P-cadherin.

ACCESSION X53614
VERSION X33614.1 GI:166
KEYWORDS cadherin; calcium binding protein; cell adhesion molecule;
transmembrane protein.

SOURCE Bos taurus (cow)

ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 2106)
AUTHORS Liaw, C.W., Cannon, C., Power, M.D., Kibonoka, P.K. and Rubin, L.L.
TITLE Identification and cloning of two species of cadherins in bovine
endothelial cells

JOURNAL EXBO J. 9 (9), 2701-2708 (1990)

MEDLINE 90360979

PUBMED 2390989

REFERENCE 2 (bases 1 to 2106)

POWER, M.D.

AUTHORS

TITLE

JOURNAL

Submitted (26-JUN-1990) Power M.D., Athena Neurosciences, 800 - F
Gateway Blvd., South San Francisco, CA 94080, USA

COMMENT Data kindly reviewed (21-NOV-1990) by Power M.

FEATURES

Location/Qualifiers
1..2106
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/cell_type="adrenal microvascular endothelial cell"
/clone_lib="lambda gt10"
c1..1478
/codon_start=3
/product="P-cadherin (AA 232-723)"
/protein_id="CAA37676.1"
/db_xref="GI:833777"
/db_xref="GOA:P19535"
/db_xref="SWISS-PROT:P19535"
/translation="ENTVSHEVRLTVTDLDPNSPAWRATYRIVGDNCDHETITTD
PESNQGLTTQKGLDFAKTHLYEVINEFVVKLPSTATVVLVSDVNEPVPV
VPSKVIIEIGESTGEPICATARDPKGSKISYHILRDPAGWLAMDPSGQVTA
GNSPQVQLNITDKLSPTAPFOALTHDSVYVTAENVKGDVALSLKFLKQGE
YDHLISLSDHGNKQLTVIRATVCDCHGNMTCDPMTWGFLLPLILGAALLLLLV
LLFLVRKKRIKEPLLIPEDTDRNPFYEGEGGEDDYDITOLHRLGARPEVVL
RNDVAPSFITFPTFRPAMPDEIGNFIENKAANDTAPPYDSLLVDFEGSGSD
AASLSLTSTSDQDQDYNILNEWGRFKLADYVGGQDD"

CDS

polyA_signal

/note="polyA signal"

polyA_site

/note="polyA site"


```

/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:8394535"
/tissue_type="Embryo, day 9 mouse (C57BL/6 background)
otocysts"
/clone_lib="NIH MGC_130"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6.1"
1..2993
/gene="Cdh3"
/notes="synonym: P-cadherin"
/db_xref="LocusID:12560"
/db_xref="MGI:88356"
1..2264
/gene="Cdh3"
/codon_start=3
/product="Cdh3 protein"
/protein_id="AAH52189.1"
/db_xref="GI:30704941"
/db_xref="LocusID:12560"
/db_xref="MGI:88356"
/translation="DIIILTRGTVOGGKDMHSPPTLRRLRRKREWMPPILFVPENGK
GFFPRLQLKSNKDRGKIFYSITIGPGADSPPEGVFTIEKESGWLHLLHMLDLREKIV
KYLEGHAVSENGASVEEPMNISIIIVQDNQNKPKFTQDITFRGSVLEGWMEGTSVMQV
TATDEDDAVNTYNGVAVSIHQEPKPHDLMTIHKSTGTSIVISSGLDREKPEYR
LTVDQADMDGEGSTTAAVQIILDANDNAPEFEQKYEAWPENEVEGHEVRLTVD
LDVPSNPAWRATYHIVGGDDGHFTITTHPTNQGLTKKGLDPEAQDQHTLYEVT
NEAPAVKLPTATAVVVHVXDVNEAPVFPVPSKVIQAEGISIGELVCITVAQDPDK
EDQKSYLISRDPAKLVADPSGOITRAGILDREDEQFVKNVYEVWVLATDSGNPP
TTGTGLTLTLTDINDHGPPEPRQIICNQSPPVQVNLITDKLSNPSFPQAQLTH
DSDIYMAEVSSEKGTDLKFLKQDITDHLSDHGRNREQLMIRATVCDCHQG
VENDCPWKGFFILPILGVALALLLALLLVKRRKVKPELLPEDTRDNVY
YEEGGEGEDQDITQLHGLREARPEVLRNDVVFTIPTMYRPREANDEIGNFI
ENLKAANTDPTAPPYDSILLVFDYEGSGDAASLSLTSASDQDQDYNVNLNWSRP
KILADWYGGGEDD"
105..389
/gene="Cdh3"
/notes="cadherin; Region: Cadherin domain"
/db_xref="CDD:pfam00028"
480..749
/gene="Cdh3"
/notes="CA; Region: Cadherin repeats. Cadherins are
glycoproteins involved in Ca2+-mediated cell-cell
adhesion. Cadherin domains occur as repeats in the
extracellular regions which are thought to mediate
cell-cell contact when bound to calcium"
/db_xref="CDD:smart00112"
768..1064
/gene="Cdh3"
/notes="cadherin; Region: Cadherin domain"
/db_xref="CDD:pfam00028"
1158..1409
/gene="Cdh3"
/notes="CA; Region: Cadherin repeats. Cadherins are
glycoproteins involved in Ca2+-mediated cell-cell
adhesion. Cadherin domains occur as repeats in the
extracellular regions which are thought to mediate
cell-cell contact when bound to calcium"
/db_xref="CDD:smart00112"
1422..1685
/gene="Cdh3"
/notes="cadherin; Region: Cadherin domain"
/db_xref="CDD:pfam00028"
1845..2252
/gene="Cdh3"
/notes="Cadherin C term; Region: Cadherin cytoplasmic
region. Cadherins are vital in cell-cell adhesion during
tissue differentiation. Cadherins are linked to the
cytoskeleton by catenins. Catenins bind to the cytoplasmic
tail of the cadherin. Cadherins cluster to form foci of
homophilic binding units. A key determinant to the
strength of the binding that it is mediated by cadherins
is the juxtaposition region of the cadherin. This region

```

```

induces clustering and also binds to the protein p120ctn"
/db_xref="CDD:pfam01049"

ORIGIN
Alignment Scores:
Pred. No.: 4,28e-43 Length: 2993
Score: 51.00 Matches: 51
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.15% Indels: 0
DB: 10 Gaps: 0

US-09-916-849A-1 (1-829) x BC052189 (1-2993)

Qy 684 LysGluProLeuLeuLeuProGluAspThrArgAspAsnValPheTyrTyrGlyGlu 703
Db 1824 AARGAGCCCTTCTGCTCCAGAGATGACACGGAGACATGCTCTTATTATGAGAA 1883
Qy 704 GluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArgGlyLeuGlu 723
Db 1884 GAGGCTGTGTGTAAGAGGACGAGGACTATGACATCACCACTCCACCGGAGCTGGAG 1943
Qy 724 AlaArgProGluValValLeuArgAspAspVal 734
Db 1944 GCCAGGCTGAGGTGGTCTCCGAACGAGTA 1976

RESULT 15
LOCUS MWCADHP 3187 bp mRNA linear ROD 30-MAR-1995
DEFINITION Mouse mRNA for P-cadherin.
ACCESSION X6340
VERSION X6340.1 GI:50267
KEYWORDS cadherin; cell adhesion molecule; glycoprotein; transmembrane
protein.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3187)
AUTHORS Nose,A., Nagafuchi,A. and Takeichi,M.
TITLE Isolation of placental cadherin cDNA: identification of a novel
gene family of cell-cell adhesion molecules
JOURNAL EXBO J. 6 (12), 3655-3661 (1987)
MEDLINE 88111554
PUBMED 3428270
FEATURES
source
1..3187
Location/Qualifiers
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="P3"
/cell_line="PSA5-E"
46..2514
/notes="unnamed protein product; pre P-cadherin (AA -27 to
795)"
/codon_start=1
/protein_id="CAA29646.1"
/db_xref="GI:50268"
/db_xref="GOA:P10287"
/db_xref="SWISS-PROT:P10287"
/translation="MELLSGFHFLLLQLQVLRWSVSPYPAGFIEAGVTLVEGV
TDLEPSOVLGKVALAGQMHADNGDIIMLTGTVOGGKDMHSPPTLRRLRRKREW
MPPIFVPEKNGKPPQSLNOLKSNKDRGKIFYSITIGPGADSPPEGVFTIEKESGWL
LHMLDREKIVKYLEGHAVSENGASVEEPMNISIIIVQDNQNKPKFTQDITFRGSVLE
GWPGTSVMQVATDEDDAVNTYNGVAVSIHQEPKPHDLMTIHKSTGTSIVISSGLD
GLDREKPEYRLTVDQADMDGEGSTTAAVQIILDANDNAPEFEQKYEAWPENEVE
GHEVRLTVDLDVPSNPAWRATYHIVGGDDGHFTITTHPTNQGLTKKGLDPEA
QDQHTLYEVTNEAPVKLPTATAVVVHVXDVNEAPVFPVPSKVIQAEGISIGEL
VCITVAQDPDKQKSYLISRDPAKLVADPSGOITRAGILDREDEQFVKNVYEVW
MVLATDSGNPTTGTGTAAGAGGACGAGGACTATGACATCACCACTCCACCGGAGCT
NSSPQAQLTHDSDIYMAEVSSEKGTDLKFLKQDITDHLSDHGRNREQLMIRATV
IRATVCDCHQVENDCPWKGFFILPILGVALALLLALLLVKRRKVKPELLPEDTRN
PEDTRDNVYEEGGEGEDQDITQLHGLREARPEVLRNDVVFTIPTMYRPREANDEI

```

PANPDEIGNFIENLPANTDPTAPPYDSLWVFDYEGSGSDAASLSLTTSASDQDQD
YNYLNEWGSRFKKLADMYGGEDD"

sig_peptide 46..126 /note="signal peptide (AA -27 to -1)"
mat_peptide 127..251 /product="P-cadherin (AA 1 - 795)"
misc_feature 619..621 /note="pot. N-linked glycosylation site"
misc_feature 1690..1692 /note="pot. N-linked glycosylation site"
misc_feature 1717..1719 /note="pot. N-linked glycosylation site"
misc_feature 1744..1746 /note="pot. N-linked glycosylation site"
misc_feature 3164..3170 /note="polyA signal"

ORIGIN

Alignment Scores:
Pred. No.: 4.54e-43 Length: 3187
Score: 51.00 Matches: 51
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.15% Indels: 0
DB: 10 Gaps: 0

US-09-916-849A-1 (1-829) x MMCADHP (1-3187)

QY 684 LysGluProLeuLeuProGluAspThrArgAspAsnValPheTyrTyrGlyGlu 703
Db 2074 AAAGAGCCCTTCTGCTCCAGAGATGACACGCGAGACAATGCTCTATTATGAGAA 2133
QY 704 GluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArgGlyLeuGlu 723
Db 2134 GAGGGTGGTGGTGAAGAGGACCAGGACTATGACATCACCACCACTCCACCGGGGACTGGAG 2193
QY 724 AlaAtGProGluValValLeuArgAsnAspVal 734
Db 2194 GCCAGCCCTGAGTGGTCTCCGAAACGATGTA 2226

Search completed: September 23, 2004, 06:21:07
Job time : 8604 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 23, 2004, 01:24:55 ; Search time 768 Seconds
(without alignments)
4585.620 Million cell updates/sec

Title: US-09-916-849A-1

Perfect score: 829

Sequence: 1 MCLPRGPIASLLLLQVCWLQ.....NEWGRFKLADMTGGGDD 829

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6745146

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-Q=/cgn2_1/USPTO_spooh_h/US0916849/runat_21092004_145834_7311/app_query.fasta_1.967
-DB=N Geneseq_29Jan04 -QFMT=fastap -SUFFIX=oligo.rng -MINMATCH=0.1 -DOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0916849 @CGN 1 1 513 @runat_21092004_145834_7311 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT_DSPELOCK=100 -LONGLOC
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -DELOP=6 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

N Geneseq_29Jan04:*
1: Geneseqm1980s:*
2: Geneseqm1990s:*
3: Geneseqm2000s:*
4: Geneseqm2001as:*
5: Geneseqm2001bs:*
6: Geneseqm2002as:*
7: Geneseqm2003as:*
8: Geneseqm2003bs:*
9: Geneseqm2003cs:*
10: Geneseqm2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	829	100.0	3171	6	Abk92214 Prostate
2	829	100.0	3171	6	Abn89393 Human P-c
3	829	100.0	3171	6	Abt07755 Breast ca
4	829	100.0	3171	6	Abk52038 DNA encod
5	829	100.0	3171	6	Abn97264 Gene #376
6	829	100.0	3171	7	ACA56750 Human sig
7	829	100.0	3171	7	Abz24736 Human P-c
8	829	100.0	3171	9	Add14795 Human src

9	829	100.0	3777	8	ACH03801
10	728	87.8	3205	6	ABQ81547
11	728	87.8	3205	7	ACC72821
12	728	87.8	3205	7	ABX76155
13	728	87.8	3205	7	ABX76399
14	715	86.2	3219	6	ADZ7637
15	102	12.3	306	5	AAH81822
16	61	7.4	2156	2	AAQ11562
17	54	6.5	255	5	AAH81872
18	52	6.3	160	7	ABZ24737
19	21	2.5	522	6	ABK53882
20	19	2.3	60	6	ABN40851
21	16	1.9	2310	7	ABZ09877
22	16	1.9	5310	9	ADE84014
23	15	1.8	2768	6	ABN89352
24	15	1.8	2768	6	ABN89353
25	15	1.8	4333	2	AAQ11563
26	15	1.8	4396	7	ABT42198
27	15	1.8	4396	9	ADB53244
28	15	1.8	4778	2	AAZ56022
29	15	1.8	4778	4	AAF86687
30	15	1.8	4828	6	ABV76385
31	15	1.8	4828	6	ABV94337
32	14	1.7	2808	2	AAT05764
33	13	1.6	317	5	ABV19790
34	13	1.6	484	5	ABV49555
35	13	1.6	2808	6	ABN89392
36	13	1.6	2815	2	AAQ5487
37	13	1.6	3048	2	AAT85433
38	13	1.6	3048	2	AAT61920
39	13	1.6	3048	6	ABK3876
40	13	1.6	3048	9	ADD14706
41	13	1.6	3310	9	ADB62391
42	12	1.4	403	8	ACH17807
43	12	1.4	436	6	ABT07156
44	12	1.4	436	6	ABX73034
45	12	1.4	467	7	ABK54052

ALIGNMENTS

RESULT 1

ABK92214
ID ABK92214 standard; DNA; 3171 BP.
XX
AC ABK92214;
XX
DT 15-AUG-2002 (first entry)
XX
DE Prostate cancer-associated DNA sequence #100.
XX
KW Prostate cancer; prostate tumour tissue; human; mammal; cytostatic;
KW gene therapy; gene; ds.
XX
OS Mammalia.
XX
XX
FN W0200203268-A2.
XX
PD 18-APR-2002.
XX
PF 12-OCT-2001; 2001WO-US032045.
XX
PR 13-OCT-2000; 2000US-00687576.
PR 08-DEC-2000; 2000US-00733288.
PR 08-DEC-2000; 2000US-00733742.
PR 24-JAN-2001; 2001US-0263957P.
PR 16-MAR-2001; 2001US-0276791P.
PR 06-APR-2001; 2001US-0276888P.
PR 24-APR-2001; 2001US-0281922P.
PR 30-APR-2001; 2001US-0286214P.
PR 04-MAY-2001; 2001US-00847046.
XX

QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
DB 114 TGGCGGCGCTCCGAGCGGTGCGGCGGTCTTCAGGAGGCTGAAGTGACCTTGGAGCG 173
QY 41 GlyValAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
DB 174 GGAGGCGCGGAGCAGGAGCGCGGCGGCGCTTGGGGAAGTATTATGGGCTGCCCTGGG 233
QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyValThr 80
DB 234 CAAGACCGAGCTCTGTTAGCACTGATATGATGATCTTCACTGTCGGAATGGCGAGACA 293
QY 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
DB 294 GTCCAGAAAGAGGTCACTGAAGAAAGAAATCCATTGAAGATCTTCCCATCAAAAGT 353
QY 101 IleLeuArgArgHisLysArgAspThrValAlaProIleSerValProGluAsnGly 120
DB 354 ATCTTACGAAAGACACAGAGAGATTGGGTGGTGTCTCCAATATCTGTCCCTGAAATGGC 413
QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgThrLys 140
DB 414 AAGGGTCCCTTCCCCAGAGACTGAATCAGCTCAAGTCTAATAAAGATAGAGACCAAG 473
QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
DB 474 ATTTTCTACAGCATCACGGGGCGGGGCGAGACGCCCCCTTGAGGGGTCTTGCTGTA 533
QY 161 GluLysGluThrGlyTrpLeuLeuLeuLysProLeuAspArgGluGluIleAlaLys 180
DB 534 GAGAGGAGACAGGCTGGTGTGTTGTAATAGCACTGGACCGGGAGAGATTGCCAAG 593
QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
DB 594 TATGAGCTCTTTGGCCACGCTGTGTGAGAGATGGTGTCTCAGTGGAGGACCCCATGAAC 653
QY 201 IleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
DB 654 ATCTCCATCATCGTACCACAGCAATGACCAAGCCCAAGTTTACCACGAGCACCTTC 713
QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
DB 714 CGAGGGAGTGTCTTAGAGGAGTCTTACCAGTACTTCTGTGATGCAGGTGACAGCCACA 773
QY 241 AspGluAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
DB 774 GATGAGGATGATGCCATCTACACCTACAAATGGGGTGTGTCTTACTCCATCCATAGCCAA 833
QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
DB 834 GAACCAAGAGACCCACACGACCTCATGTTCACAATTCACGGAGCACAGGCACCATCAGC 893
QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
DB 894 GTCATCTCCAGTGGCTGTGACCGGAAAGTCCCTGTGATGATGAGTGGAGATCCTTGAT 953
QY 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluIleLeuAsp 320
DB 954 ACAGACATGGATGGGACCGGCTCCACCACCGAGTGGCGAGTGGAGATCCTTGAT 1013
QY 321 AlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
DB 1014 GCCAATGACAATGCTCCCATGTTTACCCCAAGAGTACGAGGCCCATGTGCTGAGAAAT 1073
QY 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
DB 1074 GAGTGGGCGCATGAGGTGAGAGGCTGACGGTCACTGATCTGGACGCCCCCACTCACCA 1133
QY 361 AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
DB 1134 GGTGGCGTGGCCACCTTATCTATGCGCGGTGACGAGGAGCAATTTTACCATCACC 1193
QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400

DB 1194 ACCACCTGAGAGCAACAGGGCACTCTGCAACACAGGAAGGTTTGGATTTTGAGGCC 1253
QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
DB 1254 AAAAACAGACACACCTGTACGTTGAAGTGACCAACAGGCCCCCTTTGTGCTGAAGCTC 1313
QY 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
DB 1314 CCAACCTCCACAGCCACCATAGTGTCTCCAGCTGGAGGATGTGAATGAGGACCTGTGTGT 1373
QY 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
DB 1374 GTCCCACTCCAAAGTGTGTGAGTCCAGGAGGCGCATCCCCACCTGGGAGGCTGTGTGT 1433
QY 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
DB 1434 GTCTACACTGCAGAGACCTGCACAGGAGAAATCAAAAGATCAGCTACCGCATCCTGAGA 1493
QY 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
DB 1494 GACCCAGCAGGGTGGCTAGCCATGGACCCAGACAGTGGGCGAGTCCACAGCTGTGGGACC 1553
QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
DB 1554 CTGACCGTGGAGTGAAGCATTTGTGAGGAACAACATCTATGAGATCATGTCTTGGCC 1613
QY 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
DB 1614 ATGGACAATGGAAGCCCTCCACCACTGGCAGCGGAACCTCTCTGTAACACTGATGTAT 1673
QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
DB 1674 GTCAACGACCATGGCCAGTCCCTGAGCCCCGTGAGTCAATCACCATCTGCAACCAAGCCCT 1733
QY 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
DB 1734 GTGGGCCACGTGTGAACATCAGGACAAGGACCTGTCTCCCCACACCTCCCTTCCAG 1793
QY 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
DB 1794 GCCCAGCTCACAGATGACTCAGACATCTACTCGACGCGAGAGGTCAACGAGGAAGTGAC 1853
QY 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
DB 1854 ACAGTGGTCTTGTCTCTGAGAAAGTCTCTGAGCAGGATACATATGACGTGCACCTTCT 1913
QY 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
DB 1914 CTGTCTGACCATGGCAACAAAGAGCAGCTGACGGTGTATCAGGGCCACTGTGTGCGACTGC 1973
QY 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyValPheIleLeuProValLeu 660
DB 1974 CATGGCCATGTGAAACCTGCCCTGGACCTCGAAAGAGGTTTCATCTCTCTGTGCTG 2033
QY 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
DB 2034 GGGGCTGTCTGGCTCTGTGTCTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2093
QY 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
DB 2094 CGAAAGATCAAGGAGGCCCT 2153
QY 701 TyrGlyGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
DB 2154 TATGGCGAGAGGGGGTGGCGAAGAGGACAGGACTATGACATCACCAGCTCCACCGA 2213
QY 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
DB 2214 GGTCTGAGGCCAGCGGAGGTGTCTCCCAATGACGTGGCGCAACCAACCATCATCCCG 2273
QY 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760

1674 GTCAAGACCATGGCCAGTCCCTGAGCCCGTCAGATCACCATCTGCCAACCAAGCCCT 1733
 QY ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
 1734 GTCCGCCAGCTGCTGAACATCAGGACAGGACCTGTCTCCCCACACCTCCCTTTCCAG 1793
 QY AlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGluGlyAsp 600
 1794 GCCAGCTCACAGATGACTCAGCATCTACTGGACGCGCAGAGGTCAACGAGGAGGTGAC 1853
 QY ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
 1854 ACAGTGGTCTTCTGCTGGAAGTTCCTGAAGCAGGATACATATGACGTGCACCTTTCT 1913
 QY LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
 1914 CTGCTGACCATGGCAACAAAGAGCAGCTGACGGTGATCAGGGCCACTGTGTGGACTGC 1973
 QY HisGlyHisValGluThrCysProGlyProTyrProTyrLysGlyGlyPheIleLeuProValLeu 660
 1974 CATGCCATGTCCAAACCTGCCCTGGACCTGGAAGGAGGTTTCATCTCCCTGCTGCTG 2033
 QY GlyAlaValLeuAlaLeuLeuPheLeuLeuLeuValLeuLeuLeuValArgLysLys 680
 2034 GGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2093
 QY ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
 2094 CGAAGATCAAGAGGCCCTCTACTCCAGAGATGACCCCGTGACCAAGCTTCTTAC 2153
 QY TyrGlyGluGluGlyGlyGlyGluGluGluGluGluGluGluGluGluGluGluGlu 720
 2154 TATGCGAAGAGGGGGTGGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2213
 QY GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
 2214 GTCTGGAGCCAGCCGAGGTGGTCTCCCAATGACGTGGCACCACCAATCATCCCG 2273
 QY ThrProMetTyrArgProArgProAlaSerProAlaSerProAlaSerProAlaSerPro 760
 2274 ACACCATGTACCGTCTAGGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2333
 QY AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
 2334 AACCTGAAGCGGCTTAACACAGCCCAAGCCAGCCCGCCCTACGACACCTCTTGGTGTTC 2393
 QY AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
 2394 GACTATGAGGCGAGCGGCTCCGACGCGGCTCCCTGAGCTCCCTCACCTCTCCGCTCC 2453
 QY AspGlnAspGlnAspTyrAspTyrLeuGluGluGluGluGluGluGluGluGluGlu 820
 2454 GACCAAGACCAAGATTACATTATCTGAACAGAGTGGGCGGCGGCTTCAAGAGCTGSCA 2513
 QY AspMetTyrGlyGlyGlyGluAspAsp 829
 2514 GACATGTACGGTGGCGGAGGAGCAGC 2540

RESULT 5
 ABN97264
 ID ABN97264 standard; DNA; 3171 BP.
 XX
 AC ABN97264;
 XX
 DT 13-AUG-2002 (first entry)
 XX
 DE Gene #3762 used to diagnose liver cancer.
 XX
 KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
 KW metastatic liver tumor; cytostatic; expression profile; disease state;
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.
 XX
 OS Homo sapiens.

XX WO200229103-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 02-OCT-2001; 2001WO-US030589.
 XX
 PR 02-OCT-2000; 2000US-0237054P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 FI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
 XX WPI; 2002-426119/45.
 DR
 XX
 PT Diagnosing and detecting the progression of liver cancer, hepatocellular
 PT carcinoma or metastatic liver tumor in a patient, involves detecting the
 PT level of expression of two or more genes in a liver tissue sample.
 XX
 PS Claim 1; SEQ ID NO 3762; 298pp; English.
 XX
 CC The invention relates to a novel method for diagnosing and detecting the
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
 CC tumor in a patient, and differentiating metastatic liver cancer from
 CC hepatocellular carcinoma in a patient, involving detecting the level of
 CC expression of two or more genes represented in ABN93503-ABN97455 in a
 CC tissue sample. The method of the invention has hepatotropic, and
 CC cytosstatic activity. The method is useful for diagnosing and detecting
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic
 CC liver carcinoma in a patient. The method is useful for identifying
 CC expression profiles which serve as useful diagnostic markers as well as
 CC markers that can be used to monitor disease states, disease progression,
 CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 3171 BP; 740 A; 903 C; 864 G; 664 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0 Length: 3171
 Score: 829.00 Matches: 829
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Gaps: 0

US-09-916-849A-1 (1-829) x ABN97264 (1-3171)
 QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln 20
 Db 54 ATGGGCTCCCTCGTGGACCTCTCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113
 QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrIleGluAla 40
 Db 114 TGGCGGCTCCGAGCGCTGCCGCGGCTCTTCAGGAGCGTGAAGTACCTTGGAGGCG 173
 QY 41 GlyGlyValaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
 Db 174 GGAGGCGCGGAGCAGAGCGCGCGGCGCGCTGGGGAAGTATTTCATGGCTGCCCTGGG 233
 QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspAspPheThrValArgAsnGlyGluThr 80
 Db 234 CAAGAGCCAGCTCTGTTTAGCACTGATAATGATGACTTCACTGTGCGAATGCGGAGCA 293
 QY 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
 Db 294 GTCCAGGAAGAAGGTCACCTGAGGAAAGGAATCCATTGAAGATCTTCCATCCAAAGCT 353
 QY 101 IleLeuArgHisLysArgAspTrpValAlaProIleSerValProGluAsnGly 120
 Db 354 ATCTTACGAACACACAGAGAGATTGGTGGTGTCTCCCAATATCTGTCCCTGAAAATGSC 413

QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
DB 414 AAGGGTCCCTCCCCAGAGACTGAATCAGCTCAAGTCTAATAAGATAGAGACACCAAG 473
QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
DB 474 ATTTCCTACAGCATCACGGGGCGGGGCGAGACAGCCCCCTGAGGGGTCTTCGCTGTA 533
QY 161 GluLysGluThrGlyTyrIleLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
DB 534 GAGAGAGGACAGGCTGGTGTGTGTGTAATAAGCCACTGGACCGGAGAGAGATTGCCAAG 593
QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
DB 594 TATGAGCTCTTTGGCCACGCTGTGTACAGAGATGGTGCTCAGTGGAGGACCCCATGAAC 653
QY 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
DB 654 ATCTCATCATCTGACCGACCGAATGACCAAGCCCAAGTTTACCAGGACACCTTC 713
QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
DB 714 CGAGGGAGTGTCTTAGAGGAGTCTTACAGGTACTTCTGTGATGCAGGTGACAGCCACA 773
QY 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
DB 774 GATGAGGATGATGCCATCTACACCTACCAATGGGGTGGTGTCTTACTCCATCCATAGCCAA 833
QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
DB 834 GAACCAAGGACCCACACGACCTCATGTTCACATTCACCGGAGCACAGCCACCATCAGC 893
QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
DB 894 GTCATCTCCAGTGGCTCGACCGGAAAGTCCCTGAGTACACACTGACCATCCATCCAGGCC 953
QY 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValClnIleLeuAsp 320
DB 954 ACAGACATGATGGGACGGCTCCACCCAGGACGAGTGGCAGTAGTGGAGATCCTTGAT 1013
QY 321 AlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
DB 1014 GCCAATGACATGCTCCCATGTTTACCCCAAGAGTACGAGGCCCATGTGCTGAGAAAT 1073
QY 341 AlaValGlyHisGluValClnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
DB 1074 GCAGTGGGCGCATGAGGTGAGAGGTGACGGTCACTGATCTGGACGCCCCCACTCACA 1133
QY 361 AlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
DB 1134 CGTGGGCGTCCACCTACCTTATCATGGCGGTGACGACGCGGACCATTTTACCATCAC 1193
QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
DB 1194 ACCCACTGAGAGCAACGAGGCCATCTGACCAACAGGAAGGGTTTGGATTTTGGAGGCC 1253
QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
DB 1254 AAAAAACGACACACCTGTACGTTGAGTGGACCAACGAGGCCCTTTTGTGCTGAGGCTC 1313
QY 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
DB 1314 CCAACCTCCACAGCCACCATAGTGTCCACGTGGAGGATGTAATGAGGCCACCTGTGTTT 1373
QY 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
DB 1374 GTCCCACTCCCAAGTCTGTGAGGTCCAGGAGGAGTCCCACTGGGAGGCTGTGTGT 1433
QY 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
DB 1434 GTCTACACTGCAGAGACCTTCACAGGAGAGATCAAAAGATCAGCTACCGCATCTGAGA 1493
QY 481 AspProAlaGlyTyrIleLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500

DB 1494 GACCCAGCAGGGTGGGTAGCCATGGACCCACAGATGGGCGAGTCAAGCTGTGGGCACC 1553
QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
DB 1554 CTCGACCGTGAAGATGAGCAGTTTGTGAGGAAACAACATCTATGAAGTCAATGGTCTGGCC 1613
QY 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
DB 1614 ATGGACAATGGAAGCCCTCCACCACTGGCAGCGGAACCCCTCTCTGCTAACACTGATTGAT 1673
QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
DB 1674 GTCAACGACCATGGCCAGTCCCTGAGCCCGTCAGATCACCATCTGCACCAACCAAGCCCT 1733
QY 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
DB 1734 GTGGCCACGCTGTGAACATCAGGACAGGACCTGTCTCCCCACACCTCCCTTCCAG 1793
QY 581 AlaGlnLeuThrAspAspSerAspIleTyrTyrThrAlaGluValAsnGluGluGlyAsp 600
DB 1794 GCCCAGCTCACAGATGACTCAGACATCTACTGACGCGCAGAGGTCAACGAGGAGGTGAC 1853
QY 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
DB 1854 ACAGTGTCTTGTCCCTGAAGAAGTTCCTGAGCAGGATACATATGACGTGCACCTTCT 1913
QY 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
DB 1914 CTGTCTGACCATGGCAACAAAGAGCAGCTCAGCGTGATCAGGGCCACTGTGTGCGACTGC 1973
QY 641 HisGlyHisValGluThrCysProGlyProTyrGlyGlyGlyPheIleLeuProValLeu 660
DB 1974 CATGGCCATCTCGAAACCTCCCTGGACCTGGAAGAGGTTTCACTCTCCCTGTGCTG 2033
QY 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuLeuValArgLysLys 680
DB 2034 GGGCTGTCTGTGCTGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2093
QY 681 ArgLysIleLysGluProLeuLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
DB 2094 CGGAAGATCAAGGAGGCCCTCTCTACTCCCAAGAGATGACACCGGTGACAAACGTCTCTAC 2153
QY 701 TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
DB 2154 TATGGCGAAGAGGGGGTGGCGAAGAGGACCAAGACTATGACATCACCCAGCTCCACCGA 2213
QY 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
DB 2214 GGTCTGAGGCCAGGCGGAGGTGGTTCTCCGCAATGACGCTGGCACCAACCATCATCCCG 2273
QY 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyValAsnPheIleGlu 760
DB 2274 ACACCATGTACCGTCTTAGCCAGCCCAACCCAGATGAATCGCAACTTTAATATGAG 2333
QY 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
DB 2334 AACCTGAAGCGCGTAAACACAGACCCCAAGCCCGCTTACGACACCTCTTGTGTGTTTC 2393
QY 781 AspTyrGlnGlySerGlySerAspAlaAlaSerLysSerSerLeuThrSerSerAlaSer 800
DB 2394 GACTATGAGGGCAGCGGCTCCGACGCGCTCCCTGAGCTCCCTCAGCTCCCTCCGCGCTCC 2453
QY 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
DB 2454 GACCAAGACCAAGATTACGATTATCTGAACAGTGGGCGAGCGCTTCAAGAGCTGGCA 2513
QY 821 AspMetTyrGlyGlyGlyGluAspAsp 829
DB 2514 GACATGTACGTGGCGGGAGGACGAC 2540
RESULT 6
ACAS6750

Db 1194 ACCACCTGAGACACAGGCGCATCTGACACCCAGAGGGTTTGGATTGAGGCC 1253
Qy LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAAACAGCACACCTGTACGTTGAGTGACCAACGAGGCCCTTTGTGCTGAAGCTC 1313
Qy 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCCACCATAGTGTCTCCAGTGGAGGATGGAATGAGGACCTGTGT 1373
Qy 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1374 GTCCACCTCCAAAGTCGTTGAGGTCCAGGAGGCATCCCACTGGGAGCCTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACATGTCAGAGACCCCTGACAGGAGATCAAAAGATCAGTACCGCATCCTGAGA 1493
Qy 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCAGCAGGCTGCTAGCATGAGACCCACAGACAGTGGCGAGGTACAGCTGTGGCACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGAGGATGAGCAGTTTGTGAGGAGAACATCATCATGAGTCAATGCTGTGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db 1614 ATGGCAATGGAGGCCCTCCACCACTGGCAGGAGACCCCTTCTGCTAACACTGATTGAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGGCCAGTCCCTGAGCCCGCTGAGATCACCATCTGCAACCAAGCCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGCGCCACGCTGCTGAACATCAGGACAGGACCTGTCTCCCAACACCTCCCTTCCAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyValAsp 600
Db 1794 GCCCAGCTCAGATGACTCAGACATCTCTGGACGGCAGAGGTCAAGAGAGAGGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGGTCTTGCTCCCTGAAGAAGTCTCTGAAGCAGGATACATATGACGTGACCTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGCTGACCATGGCCAAACAGAGCAGCTGACGGTGATCAGGCGCACCTGTGTGACATGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTyrPheLysGlyPheIleLeuProValLeu 660
Db 1974 CATGGCCATGTGCAACCTGCGCCCTGAGCCCTGGAAGGAGGTTTCATCTCCCTGTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuPheLeuLeuValLeuLeuValLeuValValArgLysLys 680
Db 2034 GGGGTGCTCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2093
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAGATCAAGAGGCCCTCTACTCCAGAGATGACACCCGTGACAAAGCTCTCTAC 2153
Qy 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCGAAGAGGGGTGGCGAGAGGACCATGATGATCATCACCGAGTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
Db 2214 GGTCTGGAGGCCAGGCCGAGGTGTTTCTCCGCAATGACGTGGCCACCAACCATCATCCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleGlu 760

Db 2274 ACACCCATGTACCTCTAGGCCAGCCACCCAGATGAATCGCACTTTAATTGAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGCGCTAACACAGACCCACAGCCCCCTACGACACCCCTCTTGGTGTTC 2393
Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GATTATGAGGCGAGCGCTCCGACGCGGTCCCTGAGTCCCTACCTTCTCCGCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGCGAGCGCTTCAAGAGCTGGCA 2513
Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 2514 GACATGTACGTGGCGGAGGACGAC 2540
RESULT 7
ABZ24736
ID ABZ24736 standard; cDNA; 3171 BP.
XX
AC ABZ24736;
XX
DT 07-APR-2003 (first entry)
XX
DE Human P-cadherin cDNA.
XX
KW P-cadherin; human; colon cancer; colorectal cancer; cytostatic;
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 54..2543
FT /*tag= a
FT /product= "Human P-cadherin"
XX
FN WO200297395-A2.
XX
PD 05-DEC-2002.
XX
PF 31-MAY-2002; 2002WO-US017109.
XX
PR 31-MAY-2001; 2001US-0294225P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Reinhard C, Klinger J, Jefferson AB, Escobedo J, Randazo F;
PI Winter J, Goodson R;
XX
DR WPI; 2003-140501/13.
DR P-PSDB; ABP58357.
PT Inhibiting migration and proliferation of P-cadherin expressing cancer
PT for treating cancer, especially digestive cancer, characterized by
PT overexpression of P-cadherin, involves administering a P-cadherin
PT antagonist.
XX
PS Disclosure; Page 11-13; 129pp; English.
XX
CC The present sequence is the nucleic acid sequence for human P-cadherin
CC (placental cadherin), a calcium-dependent cellular adhesion protein. The
CC invention provides methods of treating or diagnosing cancers involving P-
CC cadherin expression using ligands that target P-cadherin, especially
CC human anti-P-cadherin antibodies. A claimed method of treating a cancer
CC characterised by the overexpression and/or upregulation of P-cadherin
CC comprises the administration of a P-cadherin antagonist, optionally
CC conjugated to a therapeutic agent. The migration, adhesion and/or
CC proliferation of the cancer is inhibited, and the method is especially
CC useful for treating or preventing a digestive cancer such as colon or
CC colorectal cancer. The antagonists may be an anti-P-cadherin antibody or
CC its fragment, a ribozyme or antisense oligonucleotide. A transgenic

QY 641 HisGlyHisValGluThrCysProGlyProThrPylsGlyGlyPheIleLeuProValleu 660
 Db 1974 CATGCCCATGTGGAACCTGCCCTGGAGGAGGAGGTTTCATCTCCCTGTGCTG 2033
 QY 661 GlyAlaValLeuAlaLeuPheLeuLeuValLeuValLeuValLeuValLeuValLeu 680
 Db 2034 GGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2093
 QY 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
 Db 2094 CGAAAGATCAAGAGCCCTCTACTCCAGAGATGACACCCGTCGACCAAGCTTCTTAC 2153
 QY 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
 Db 2154 TATGCCGAGAGGGGGTGGCGAAGAGGACCGAGCTATGACATCACCCAGCTCCACGA 2213
 QY 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
 Db 2214 GGTCTGGAGGCGGAGGCTGCTTCTCCGCAATGACGTGGCACCACCAATCATCCCG 2273
 QY 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleLeuGlu 760
 Db 2274 ACACCCATGACCGTCCCTAGGCCACCAACCCAGATGAATGCGGCACTTTATATTGAG 2333
 QY 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
 Db 2334 AACCTGMAAGCGGTAAACACAGACCCACAGCCCGCTACGACACCCCTCTTGTGTTC 2393
 QY 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
 Db 2394 GACTATGAGGGCAGCGGCTCCGACGCGGCTCCCTGAGCTCCCTCACCTCTCCGCTCC 2453
 QY 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLeuAla 820
 Db 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGCGCTTCAAGAAGCTGCA 2513
 QY 821 AspMetTyrGlyGlyGlyGluAspAsp 829
 Db 2514 GACATGTACGTGGTGGCGGAGGACGAC 2540
 RESULT 8
 ADD14795
 ID ADD14795 standard; cDNA; 3171 BP.
 AC ADD14795;
 XX AC
 XX AC
 DT 01-JAN-2004 (first entry)
 XX DT
 XX DE Human src biomarker polynucleotide SEQ ID NO:189.
 XX DE
 KW predictor set; protein tyrosine kinase activity modulator;
 KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;
 KW gene therapy; drug sensitivity; genetic profile; cancer; human; gene; ss.
 XX KW
 XX OS Homo sapiens.
 XX OS
 PN WO2003062395-A2.
 XX PN
 XX XX
 XX PD 31-JUL-2003.
 XX PD
 XX PF 17-JAN-2003; 2003WO-US001981.
 XX PF
 XX PR 18-JAN-2002; 2002US-0350061P.
 XX PR
 XX XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX PA
 XX XX Huang F, Fairchild CR, Lee FY, Shaw P;
 XX PI
 XX WPI; 2003-636735/60.
 XX DR
 XX DR P-PSDB; ADD14190.
 XX PT
 PT New polynucleotides and polypeptides for predicting the activity of
 PT compounds that interact with protein tyrosine kinases and/or protein

PT tyrosine kinase pathways.
 XX Claim 2; SEQ ID NO 189; 139pp; English.
 XX
 CC The present invention describes a predictor set comprising a plurality of
 CC polynucleotides or polypeptides whose expression pattern is predictive of
 CC the response of cells to treatment with a compound that modulates protein
 CC tyrosine kinase activity or members of the protein tyrosine kinase
 CC pathway. Also described: (1) predicting whether a compound is capable of
 CC modulating the activity of cells, comprising obtaining a sample of cells,
 CC determining whether the cells express a plurality of markers, and
 CC correlating the expression of the markers to the compound's ability to
 CC modulate the activity of the cells; (2) a plurality of cell lines for
 CC identifying polynucleotides and polypeptides whose expression levels
 CC correlate with compound sensitivity or resistance of cells associated
 CC with a disease state; and (3) identifying polynucleotides and
 CC polypeptides that predict compound sensitivity or resistance of cells
 CC associated with a disease state, comprising subjecting the plurality of
 CC cell lines to one or more compounds, analysing the expression pattern of
 CC a microarray of polynucleotides or polypeptides, and selecting
 CC polynucleotides or polypeptides that predict the sensitivity or
 CC resistance of cells associated with a disease state by using the
 CC expression pattern of the microarray. The polynucleotides and
 CC polypeptides have cytostatic activities, and can be used in gene therapy.
 CC The polynucleotides and polypeptides are useful in predicting the
 CC activity of compounds that interact with protein tyrosine kinases and/or
 CC protein tyrosine kinase pathways. These may be used in determining drug
 CC sensitivity in patients to allow the development of individualized
 CC genetic profiles which aid in treating diseases and disorders (e.g.
 CC cancer) based on patient response at a molecular level. The present
 CC sequence is used in the exemplification of the present invention.
 XX
 SQ Sequence 3171 BP; 740 A; 903 C; 864 G; 664 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0 Length: 3171
 Score: 829.00 Matches: 829
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0
 US-09-916-849A-1 (1-829) x ADD14795 (1-3171)
 QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuLeuValCysTrpLeuGln 20
 Db 54 ATGGGGCTCCCTCGTGGACCTCTCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113
 QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
 Db 114 TGGCGGCTCCGAGCGCTGCGGCGGCTCTTTCAGGGAGGCTGAGTGACCTTGGAGCG 173
 QY 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
 Db 174 GGAGCGCGGAGCAGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 233
 QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspAspPheThrValArgAsnGlyGluThr 80
 Db 234 CAAGAGCCAGCTCTGTATTAGCATTGATGATGATGATGATGATGATGATGATGATGAT 293
 QY 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
 Db 294 GTCCAGGAAAGAGGTCACCTGAAGAAAGAAATCCATTGAAGATCTTCCCATCAACAGT 353
 QY 101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
 Db 354 ATCTTACGAAAGACACAAAGAGAGATTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 413
 QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
 Db 414 AAGGTCCTCTCCCGCAGAGACTGATCAGCTCAAGTCTAATAAGATAGACACCAAG 473
 QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160

Db 474 ATTTTCTACAGCATACGGGGCGGGGCGAGACAGCCCTTGGTGTCTTCCTGTA 533
 QY 161 GluLysGluThrGlyTrrpLeuLeuAsnLysProLeuAspArgGluLileAlaLys 180
 Db 534 GAGAAGGAGACAGGCTGGTGTGTGTAATAAGCCACTGGACCGGAGAGATTCCCAAG 593
 QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
 Db 594 TATGACCTCTTTGGCCACGCTGTGTGAGAGATGGTGCTCAGTGGAGGACCCATGAAC 653
 QY 201 IleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
 Db 654 ATCTCCATCATGCTGACCGACAGATGACCAAGCCCAAGCTTTACCCAGGACACCTTC 713
 QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
 Db 714 CAGGGAGTGTCTTAGGGAGTCTTACCAAGTACTTCTGTGATGCGAGGTGACAGCCACA 773
 QY 241 AspGluAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
 Db 774 GATGAGGATGATGCCATCACACCTACAAATGGGGTGGTGTCTTACTCCATCCATGCCAA 833
 QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
 Db 834 GAACCAAGGACCCACACACCTCATGTTCACAAATTCACGGGACAGGACCATCAGC 893
 QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
 Db 894 GTCATCTCCAGTGGCTGGACCGGAAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
 QY 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluIleLeuAsp 320
 Db 954 ACAGACATGGATGGGACCGCTCCACCCACCGGAGTGGCAGTGTGAGATCCCTTGAT 1013
 QY 321 AlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
 Db 1014 GCCAATGACATGCTCCCATGTTGNCCCCAGAGATACGAGGCCCATGTGCCTGAGAT 1073
 QY 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
 Db 1074 GCAGTGGGCGCATGAGTGCGAGAGGTGACGGTCACTGATCTGCGACGCCCAACTCACCA 1133
 QY 361 AlaTrrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
 Db 1134 GGTGGCGTGCACCTTACCTTATCATGGCGGTGACGACGGGACATTTTACCATCACC 1193
 QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
 Db 1194 ACCCACCTTGAGAGAACACAGGGCATCTCTGACACAGAGAGGGTTCGATTTTGAGGCC 1253
 QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
 Db 1254 AAAAACCCAGCACACCTGTACCTTGAAGTGAACACGAGGCCCTTTTGTGTGTAAGCTC 1313
 QY 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
 Db 1314 CCAACCTCCACAGCCACCATATGTTGTCACGTGGAGGATGTGAATGAGGCCACCTGTGTGT 1373
 QY 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
 Db 1374 GTCCCAACCTCCAAAGTGTGTAGTGTCCAGGAGGCGATCCCCACTGGGGAGCCCTGTGTGT 1433
 QY 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
 Db 1434 GTCTACACTGCGAGAGACCTTGACAGGAGAGATCAAAAGATCAGTACCGCATCTCTGAGA 1493
 QY 481 AspProAlaGlyTrrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
 Db 1494 GACCCAGCAGGGTGGCTAGCCATGACCCAGACAGTGGGCGAGGTCCACAGCTGTGGGCACC 1553
 QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520

Db 1554 CTCGACCGTGAGGATGAGCAGTGTGTGAGGAACAACATCTATGAAGTCATGTGTTGGCC 1613
 QY 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeuAsp 540
 Db 1614 ATGGACATATGGAGCCCTCCACACTGGCACGGGAACCTTCTGTCTAACACTATTGAT 1673
 QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
 Db 1674 GTCAACGACCATGCCCGCAGTCCCTGAGCCCGTCAGATCACCATCTGCAACCAAGCCCT 1733
 QY 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
 Db 1734 GTGCGCCACCTGTCTGAACATCAGGACAAGGACTGTCTCTCCACACCTCTCCCTTCCAG 1793
 QY 581 AlaGlnLeuThrAspAspSerAspIleTyrTrrpThrAlaGluValAsnGluGlyAsp 600
 Db 1794 GCCAGCTCACAGATGACTCAGACATCTACTGGACGCGAGAGGTCAACGAGGAAGTGAC 1853
 QY 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
 Db 1854 ACAGTGTCTGTCTCCCTGAAGAAAGTTCCTGAAGCGGATACATATGACGTGCACCTTCT 1913
 QY 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
 Db 1914 CTGTCTGACCATGCGCAACAAGAGCAGCTGACGTGATCAGGGCCACTGTGTGGACTGC 1973
 QY 641 HisGlyHisValGluThrCysProGlyProTrrpLysGlyGlyPheIleLeuProValLeu 660
 Db 1974 CATGGCCATCTCGAAACCTGCCCTGGACCTGAAAGAGGTTTCATCTCCTCTGTGCTG 2033
 QY 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
 Db 2034 GGGCTGTCTCTGTCT 2093
 QY 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
 Db 2094 CGGAAGATCAAGGAGCCCT 2153
 QY 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnIleuHisArg 720
 Db 2154 TATGGCGAAGAGGGGGTGGCGAAGAGGACCAAGGACTATGACATCACCAGCTCCACCGA 2213
 QY 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
 Db 2214 GGTCTGGAGCCAGCGCGAGGTGGTCTCTCCGATGACGTGGCACCACCATCATCCCG 2273
 QY 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
 Db 2274 ACACCATGTACCGTCTTAGGCCAGCAACCCAGATGAAATCGCAACTTTTATAATTGAG 2333
 QY 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
 Db 2334 AACCTGAAGCGCTAACACAGACCCACAGCCCGCCCTACGACACCTCTTGTGTGTTTC 2393
 QY 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerIleuThrSerSerAlaSer 800
 Db 2394 GACTATGAGGCGAGCGGCTCCGACGCGCGTCCCTGAGTCCCTCCTCCTCCTCCTCCTCC 2453
 QY 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrrpGlySerArgPheLysLysLeuAla 820
 Db 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCGCTTCAAGAGCTGGCA 2513
 QY 821 AspMetTyrGlyGlyGluAspAsp 829
 Db 2514 GACATGTACGTGGCGGAGGAGCAG 2540

RESULT 9

ID ACH03801 standard; cDNA; 3777 BP.
 XX
 AC ACH03801;

XX
 DT 26-SEP-2003 (first entry)

Db 131 TGGCGGGCTCCGAGCCGTGCGGGCGGTCTTACGAGGAGCTGAAGTGAACCTTGAGGCG 190
Qy 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db 191 GGAGCGCGAGCAGAGCCCGCCAGCGCTGGGGAAGTATTCATGGCTGCGCTGGG 250
Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
Db 251 CAAGAGCCAGCTCTGTTAGCACTGATAATGATGACTTCACTGTGCGGAATGGCGAGACA 310
Qy 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
Db 311 GTCCAGGAAAGAGTCACTGAAGAAAGGAATCCATTGAAGATCTTCCCATCCAAACGT 370
Qy 101 IleLeuArgArgHisLysArgAspTrpValAlaAlaProIleSerValProGluAsnGly 120
Db 371 ATCTTACGAGACACAAGAGATTGGGTGGTGTCTCCAAATATCTGTCTGAAATGGC 430
Qy 121 LysGlyProPheProGluArgLeuAsnGlnIleLysSerAsnLysAspArgAspThrLys 140
Db 431 AAGGTCCTCTTCCCCCAGAGACTGAATCAGCTCAAGTCTAATGAAGATAGACACCAAG 490
Qy 141 IlePheTyrSerIleThrGlyProGlyValAlaAspSerProGluGlyValPheAlaVal 160
Db 491 ATTTTCTACAGCATCACGGGCGCGGGGCGAGACAGCCCCCTGAGGGHGTCTTCGTGTA 550
Qy 161 GluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
Db 551 GAGAAGGAGACAGGCTGCTGTTGTTGAATAAGCCACTGGACCCGGAGAGATTGCCAAG 610
Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 611 TATGAGCTCTTTGGCCACGCTGTGTGAGAGAATGGTCTCAGTGGAGGACCCCATGAAC 670
Qy 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 671 ATCTCCATCATCTGACCGACAGAGTACCAAGCCCAAGTTTACCCAGACACCTTC 730
Qy 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 731 CGAGGAGTGTCTTAGAGGAGTCTTACCAGGTACTTCTGTGATCGAGGTGACAGCCACG 790
Qy 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
Db 791 GATGAGGATGATGCCATCTACCTACAAATGGGTGGTGTGTTACTTCATCATAGCCAA 850
Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 851 GAACCAAGGACCCACACAGACTCATGTTTCCATTACCCGAGCAGAGCACCATCAGC 910
Qy 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db 911 GTCATCTCCAGTGGCTGGACCGGAAAAGTCCCTGAGTACACACTGACCATCCAGGCC 970
Qy 301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValAlaValGluIleLeuAsp 320
Db 971 ACAGACATGGATGGGACGGCTCCACCACCGCAGTGGCAGTAGTGGAGATCCTTGAT 1030
Qy 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1031 GCCAATGAATGCTCCCATGTTTGAACCCCGAGAAGTACGAGGCCCATGTGCTCTGAGAAT 1090
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1091 GCAGTGGCCCATGAGTGCAGAGGCTGACGCTACTGATCTGGACGCCCCCACTCACCA 1150
Qy 361 AlaTrpArgAlaThrTyrLeuIleMetGlyLysAspAspGlyAspHisPheThrIleThr 380
Db 1151 GCGTGGCTGTCACCTACTTATCATGGCGGTGACGACGGGACCATTTTACCATCACC 1210
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1211 ACCCACCTTGAGCAACAGGCAATCTTGACAAACAGGAAGGTTTGGATTTTGGAGGCC 1270

Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1271 AAAAACAGACACACCTGTACGTTGAAGTGACCAACAGAGCCCTTTTGTGCTGAAGCTC 1330
Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1331 CCAACCTCCACAGCCACCATAGTGGTCCACGTGGAGGATGTGAATGAGGCACCTGTGTTT 1390
Qy 441 ValProProSerLysValValGluValGlnGluGlyIleProThrThrGlyGluProValCys 460
Db 1391 GTCCCACTCCCAAAGTCGTTGAGGTCACAGAGGGCATCCCCACTGGGAGCCCTGTGTGT 1450
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1451 GTCTACACTGCAGAGACCCCTGCACAAGAGAAATCAAAAGATCAGCTACCCGACTCCTGAGA 1510
Qy 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1511 GACCCAGCAGGGTGGGTAGCCATGSCACCCAGACAGTGGGCGAGGTACAGCTGTGGGCACC 1570
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1571 CTGCACGTGAGGATGAGCAGTTTGTGAGGAACACATCTATGAAGTCACTGTCTTGGCC 1630
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db 1631 ATGACAATGGAAGCCCTCCACCACTGGCACGGGAACCTTCTGCTAAACACTGATTGAT 1690
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1691 GTCAATGACCATGGCCAGTCCCTGAGCCCGCTCAGATCACCATCTGCACCAACCAACCCCT 1750
Qy 561 ValArgHis-ValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1751 GTGCGCCA-GGTCTGAACATCACGGACAAGGACCTGTCTCCCAACACCTCCCTTTCCA 1809
Qy 580 nAlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGluLys 600
Db 1810 GGCCCCAGCTCACAGATGACTCAGACATCTACTGCGCGCAGAGGTCAACAGAGAAAGTGA 1869
Qy 600 pThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSe 620
Db 1870 CACAGTGGTCTTGCTCCCTGAAGAAGTTCCTGAAGCAGGATACATATGACGTGCACCTTC 1929
Qy 620 rLeuSerAspHisGlyAsnLysGluGlnIleuThrValIleArgAlaThrValCysAspCy 640
Db 1930 TCTGTCTGACCATGGCAACAAGAGAGCAGCTGACGGTGAATCAGGGCCACTGTGTGCGACTG 1989
Qy 640 sHisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleuProValle 660
Db 1990 CCAATGCCATGTCGAAACCTGCTCCCTGACCCCTGGAAGGAGGTTTCATCTCTCCTGTGCT 2049
Qy 660 uGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValValArgLysLys 680
Db 2050 GGGGGCTGTCTGGCTCTGCTGTCTCTCTGCTGGTGTGCTTTCTGGTGAGAAAGAA 2109
Qy 680 sArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTy 700
Db 2110 GCGGAAGATCAAGAGCCCTCTCTACTCCAGAAGATGACACCCGTCGACACGCTTCTTA 2169
Qy 700 rTyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisAr 720
Db 2170 CTATGGCGAAGAGGGGGTGGCGAAGAGGACCCAGGACTATGACATCACCCAGCTCCACCG 2229
Qy 720 gGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePr 740
Db 2230 AGGTCTGGAGGCCAGGCCGAGGTGTTCTCCGCAATGACGTGGCAACCAACCATCATCCC 2289
Qy 740 oThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleG 760
Db 2290 GACACCCATGTACCGTCTCGGCCAGCCACCCAGATGAAATCGGCAACTTTATATATGA 2349

QY 760 uAsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPh 780
 DB 2350 GAACCTGAAGCGCGTAAACACAGACCCACAGCCCGCCCTACGACACCCCTCTTGGGTGT 2409
 QY 780 eAspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSe 800
 DB 2410 CGACTATGAGGCGAGCGGCTCGACGCGCGTCCCTGAGCTCCCTCACCTCTCCGCGCTC 2459
 QY 800 rAspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLysLeuAl 820
 DB 2470 CGACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCGCTTCAAGAAAGCTGGC 2529
 QY 820 aAspMetTyrGlyGlyGlyGluAspAsp 829
 DB 2530 AGACATGTACGTTGCGGGGAGCAGAC 2557

RESULT 11

ACC72821
 ID ACC72821 standard; cDNA; 3205 BP.

XX AC ACC72821;

XX 09-JUL-2003 (first entry)

XX Human cancer related protein encoding cDNA SEQ ID NO:159.

XX Human; cancer; diagnosis; screening; modulator; leukaemia; ischaemia;
 KW heart disease; atherosclerosis; endometriosis; gene; ss.

XX Homo sapiens.

XX WO2003025138-A2.

XX 27-MAR-2003.

XX 17-SEP-2002; 2002WO-US029560.

XX 17-SEP-2001; 2001US-0323469P.

XX 20-SEP-2001; 2001US-0323887P.

XX 13-NOV-2001; 2001US-0350666P.

XX 08-FEB-2002; 2002US-0355145P.

XX 08-FEB-2002; 2002US-0355257P.

XX 12-APR-2002; 2002US-0372246P.

XX (E089-) EOS BIOTECHNOLOGY INC.

XX Afar D, Aziz N, Gish KC, Hevezi PA, Mack DH, Wilson KE;

XX Zlotnik A;

XX WPI; 2003-354600/33.

XX P-PSDB; ABR58670.

XX New genes that are up-regulated or down-regulated in cancers, useful as
 PT markers for diagnosing e.g. cancer, ischemia or heart diseases, or as
 PT therapeutic targets for screening drugs for treating these diseases.

XX Claim 8; Page 715; 767pp; English.

XX The present invention describes an isolated nucleic acid molecule, which
 CC comprises the sequence of any of the genes that are up-regulated or down-
 CC regulated in specific cancers (e.g. about 1031 genes up-regulated in
 CC acute lymphocytic leukemia). ACC72641 to ACC72860 represent cancer
 CC related gene nucleotide sequences which encode the proteins given in
 CC ABR5821 to ABR58709. Also described: (1) determining the presence or
 CC absence of a pathological cell in a patient; (2) an expression vector
 CC comprising a nucleic acid molecule described above; (3) a host cell
 CC comprising the vector; (4) an isolated polypeptide, which is encoded by
 CC the nucleic acid; (5) an antibody that specifically binds the polypeptide
 CC of (4); (6) specifically targeting a compound to a pathological cell in a
 CC patient by administering to the patient the antibody above; and (7) a
 CC drug screening assay. The nucleic acid is useful as diagnostic markers or
 CC therapeutic targets. In particular, the nucleic acid is useful for
 CC diagnosing a pathology, e.g. cancer (e.g. cancer of the bone marrow,

CC bladder, brain, breast, cervix, colon/rectum, kidney, lung, ovary,
 CC pancreas, prostate, skin and uterus), wounds, ischaemia, heart diseases,
 CC atherosclerosis and endometriosis. The nucleic acid is also useful in
 CC drug screening, particularly for identifying agents for treating these
 CC pathologies
 XX
 SQ Sequence 3205 BP; 751 A; 909 C; 875 G; 670 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 0 Length: 3205
 Score: 728.00 Matches: 828
 Percent Similarity: 99.76% Conservative: 0
 Best Local Similarity: 99.76% Mismatches: 1
 Query Match: 87.82% Indels: 2
 DB: 7 Gaps: 0

US-09-916-849A-1 (1-829) x ACC72821 (1-3205)

QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln 20
 DB 71 ATGGGGCTCCCTCGTGGACCTCTCGCGTCTCTCCCTCTCTCCAGGTTTGGCTGGCAG 130
 QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
 DB 131 TGGCGGGCTCCGAGCGGTCGCGGGGCTTTCAGGGAGGCTGAAGTGACCTTGGAGGCG 190
 QY 41 GlyGlyValaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
 DB 191 GGAGGCGCGAGCAGGAGCGCGCCAGCGCTGGGGAGAGTATTTCATGGCTGCCCTGGG 250
 QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspAspPheThrValArgAsnGlyGluThr 80
 DB 251 CAAGAGCCAGCTCTGTTTAGCCTGATAATGATGACTTCACCTGCGGGAATGCCAGACA 310
 QY 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
 DB 311 GTCCAGGAAGAAGGTCACCTGAAGGAAGAAGTAATCCATTCAAGATCTTCCATCCAAAG 370
 QY 101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
 DB 371 ATCTTACGAAGACACAGAAGAGATTGGGTGGTGTCTCAATATCTGCTCCCTGAATGGC 430
 QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
 DB 431 AAGGTTCCTTCCCCCAGAGACTGAATCAGCTCAAGCTTAATAAGATAGACACACCAAG 490
 QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
 DB 491 ATTTTCTACAGCATCACGGGGCGGGGGCAGACAGCCCCCTCGAGGGTGTCTTCGCTGTA 550
 QY 161 GluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluLulleAlaLys 180
 DB 551 GAGAAGAGACAGCGCTGGTGTGTTTGAATAGCCACTGACCGGGAGAGAGATTGCCAAG 610
 QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
 DB 611 TATGAGCTCTTTGGCCACGCTGTGTGAGAGAATGGTCCCTCAGTGAGGAGACCCCATGAAC 670
 QY 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
 DB 671 ATCTCCATCATGCTGACCCAGCAGAGATGACCAAGCCCAAGTTTACCAGAGACCTTTC 730
 QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
 DB 731 CGAGGAGAGTGTCTTAGGGAGGTCTTACAGGTACTTCTGTGATGCGAGTGCAGCAGCG 790
 QY 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
 DB 791 GATGAGGATGATGCCATCTACACCTACAATGGGGTGGTGTCTTACTTCCATCCATAGCCA 850
 QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
 DB 851 GACACCAAGAGACCCACACGACCTCATGTTCCATTCACCGGAGCAGACGACCATCAGC 910

QY 281 ValIleSerSerGlyLeuAspArgGluysValProGluTyrThrLeuThrIleGlnAla 300
Db 911 GTCACTCTCCAGTGGCTGGACCGGAAAAAGTCCCTGAGTACACACTACCATCCAGGCC 970
QY 301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValAlaValGluIleuAsp 320
Db 971 ACAGACATGGATGGGGAAGCTCCACCCAGGAGTGGAGTAGTGAGATCCCTGGAT 1030
QY 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1031 GCCAATGACAATGCTCCCATGTTTACCCCGGAGAGTACGAGGCCCATGTGCTGAGAA 1090
QY 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1091 GCAGTGGGCCATGAGGTGAGAGGCTGACGGTCACTGATCGGACGCCCCCAACTCACCA 1150
QY 361 AlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db 1151 GGTGCGCTGCCACCTACCTATCATGCGCGGTGACGAGCGGGACCATTTTACCATCAC 1210
QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1211 ACCCACTCGAGAGAACCGGCCATCTCTGACCAACAGAGAGGGTGTGGATTTTGAGGCC 1270
QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1271 AAAACCGAGCACACCTGTGAGTGAAGTGAACACAGAGAGGCCCTTTGTGCTGAGCTC 1330
QY 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
Db 1331 CCAACCTCCACAGCCACCATAGTGTGTCACGTGGAGGATGTGAATGAGGCACCTGTGTGT 1390
QY 441 ValProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1391 GTCCACCCCTCCAAAGTCGTGAGTCCAGAGGGCATCCCACTGGGGAGCCCTGTGTGT 1450
QY 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1451 GTCTACACTGCAGAGAGCCCTGCACAGGAGATCAAAAGATCAGCTACCGCATCTTGAGA 1510
QY 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1511 GACCCAGAGGGTGGCTAGCCATGACACCCAGACAGTGGGACAGTGCACAGCTGTGGCACC 1570
QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1571 CTCGACCGTGGAGTATGACAGCTTTGTGAGAGAACACATCTATGAGTCACTGGCTTGGCC 1630
QY 521 MetAspAsnGlySerProProThrThrThrGlyThrLeuLeuLeuLeuLeuLeuLeuLeu 540
Db 1631 ATGGACAATGGAAGCCCTCCACCACTGGCACGGGAACCCCTTCTGCTAAACACTGATTGAT 1690
QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1691 GTCATGACCATGGCCCACTCCCTGAGCCCGTCCAGATCACCATCTGCAACCAAGGCCCT 1750
QY 561 ValArgHis-ValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1751 GTGGGCCA-GGTGCTGAACATACCGACAGACAGCCCTGTCTCCACACCTCCCTTTTCCA 1809
QY 580 nAlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGluGlyAs 600
Db 1810 GGCCAGCTCAGATGACTCAGACATCTACTTGACCGCAGAGGTCAACGAGGAAGTGA 1869
QY 600 pThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHisLeuSe 620
Db 1870 CACAGTGTCTGTCTCTGAGAGAGTCTCTGACGAGGATACATATGACGTGCACCTTTC 1929
QY 620 rLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1930 TCTGTCTGACCATGCGCAACAAAGAGCAGCTGACGGTGTATCAGGGCCACTGTGTGCGACTG 1989

QY 640 sHisGlyHisValGluThrCysProGlyProGlyProGlyLysGlyGlyPheIleLeuProValle 660
Db 1990 CCATGGCCATGTGCAAAACCTGCGCTGACCTGGAAGGGAGGTTTCATCTCCCTGTGCT 2049
QY 660 uGlyAlaValLeuAlaLeuLeuPheLeuLeuLeuValLeuLeuLeuValArgLysLeu 680
Db 2050 GGGGCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2109
QY 680 sArgLysIleLysGlyProLeuLeuLeuProGluAspAspThrArgAspAsnValPhe 700
Db 2110 GCGGAAGATCAAGGAGGCCCTCTACTCCAGAAAGATGACACCCGTGACAACTCTTCTA 2169
QY 700 rTyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisAr 720
Db 2170 CTATGGCGAAGAGGGGGGTGGCGAAGGAGGACCAAGGACTATGACATCACCAGCTCCACCG 2229
QY 720 gGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePr 740
Db 2230 AGTCTGGAGGCCAGCGCGGAGGTGTTCTCCGCANTGACGTGGCACCACCATCATCCC 2289
QY 740 oThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleG 760
Db 2290 GACACCATGTACCGCTCTCGCCAGCAACCCAGATGAATCGCAACTTTATAATTGA 2349
QY 760 uAsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPh 780
Db 2350 GAACCTGAAGCGCGCTTAACACAGACCCCGCCCGCCCTAGCACACCTCTTGTGTT 2409
QY 780 eAspTyrGluGlySerGlySerAspAlaLaserLeuSerSerLeuThrSerSerAlase 800
Db 2410 CGACTATGAGGGCAGCGGCTCCGACCGCGTCCCTGAGCTCCCTCACCTCTCCGCTC 2469
QY 800 rAspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLeuAl 820
Db 2470 CGACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCCGCTTCAAGAAGCTGGC 2529
QY 820 aAspMetTyrGlyGlyGlyLysAsp 829
Db 2530 AGACATGTACGTGGTGGGGGAGGAGCGAC 2557
RESULT 12
ABX76155
ID ABX76155 standard; DNA; 3205 BP.
XX AC ABX76155;
XX DT 02-APR-2003 (first entry)
XX DE Lung cancer-associated polynucleotide #27.
XX KW Lung cancer-associated polynucleotide; gene; ds; cytostatic; emphysema;
XX KW antiinflammatory; antiaschmatic; non-small cell lung cancer; atelectasis;
XX KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
XX KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
XX KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
OS Unidentified.
XX WO200286443-A2.
XX PN WO200286443-A2.
XX PD 31-OCT-2002.
XX PF 18-APR-2002; 2002WO-US012476.
XX PR 18-APR-2001; 2001US-0284770P.
XX PR 10-MAY-2001; 2001US-0290492P.
XX PR 09-NOV-2001; 2001US-0339245P.
XX PR 13-NOV-2001; 2001US-0350666P.
XX PR 29-NOV-2001; 2001US-0334370P.
XX PR 12-APR-2002; 2002US-0372246P.
XX PA (BOSB-) EOS BIOTECHNOLOGY INC.
XX

Qy 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
Db 131 TCGCGGGCTCCGAGCGGTGCGGGGGTCTTCAGGGAGGTGAAGTACCTTGGAGGCG 190
Qy 41 GlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db 191 GGAGCGCGGACGAGAGCCCGCGCGAGCGCTGGGGAAGTATTATCGGGCTGCCCTGGG 250
Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
Db 251 CAAGAGCCAGCTCTGTTAGCACTGATATGATGACTTCACTGCGGAATGGCGAGACA 310
Qy 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerIleArg 100
Db 311 GTCCAGGAAGAAGGTCACTGAAGGAAGGAATCCATTGAAGATCTTCCCATCCAAACGT 370
Qy 101 IleLeuArgArgHisLysArgAspThrValAlaLapProIleSerValProGluAsnGly 120
Db 371 ATCTTACGAAGACACAAGAGAGATTGGGTGGTGTCTCCAAATATCTGTCCTGAAAAATGGC 430
Qy 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgThrLys 140
Db 431 AAGGGTCCCTTCCCCAGAGACTGAATCAGCTCAAGTCTAATAAGATAGAGACACCAAG 490
Qy 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
Db 491 ATTTTCTACAGCATCACGGGGCGGGGGCAGACAGCCCCCTGAGGGTGTCTTCGCTGTA 550
Qy 161 GluLysGluThrGlyTyrLeuLeuLeuLysProLeuAspArgGluGluIleAlaLys 180
Db 551 GAGAGGAGACAGGCTGGTGTGTGTTGTAATAGCCACTTGGACGGAGAGATTGCCAAG 610
Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 611 TATGAGCTCTTTGGCCACGCTGTGTGAGAGATGGTGTCTCAGTGGAGGACCCCATGAAC 670
Qy 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 671 ATCTCCATCATCGTACGCCACGACGATGACCAAGCCCAAGTTTACCAGGACACCTTC 730
Qy 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 731 CGAGGGAGGTGCTTAGAGGGAGTCTTACCGAGTACTTCTGTGATCGAGGTGACAGCCACG 790
Qy 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerCln 260
Db 791 GATGAGGATGATGCCATCTACACCTTACAAATGGGGTGGTGTGCTTACTCCATCCATAGCCAA 850
Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 851 GAACCAAGGACCCACACGCTCATGTTCACCATTCACCGGAGCACAGGCACCATCAGC 910
Qy 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db 911 GTCATCTCCAGTGGCTGTGACCGGGAAGTCCCTGTGATCACACTGACCATCCAGGCC 970
Qy 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValAlaValGluIleLeuAsp 320
Db 971 ACAGACATGGATGGGACCGGCTCCACCACCGGAGTGGCAGTAGTGGAGATCCTTGAT 1030
Qy 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1031 GCCAATGCAATGCTCCCATGTTTACCCCAAGAGTACGAGGCCCATGTGCTGAGAAT 1090
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1091 CGAGTGGGCGCATGAGGTGAGAGGTGACGGTCACTGATCTGGACGCCCCCACTCACA 1150
Qy 361 AlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db 1151 GCGTGGCGTGCCACCTTACCTTATCATGGCGGTGACGCGGGACCAATTTTACCATCACC 1210

Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1211 ACCCACCTTGAGAGCAACAGGCGCATCTTCAACAACAGGAAGGTTTGGATTTTGAGGCC 1270
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1271 AAAAACAGACACACCTGTACGTTGAAGTGACACAGAGGCCCTTTGTGCTGAAGCTC 1330
Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1331 CCAACCTCCACAGCCACCATAGTGTGTCCACGTGGAGATGTGAATGAGGACCTGTGTGT 1390
Qy 441 ValProProSerLysValValGluValGlnGluGlyIleProThrClyGluProValCys 460
Db 1391 GTCCCCACCTTCCAAAGTCTGTGAGGTCCAGGAGGGCATCCCCACTGGGAGGCTGTGTGT 1450
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1451 GTCTACACTGCAGAACCCCTGCACAGGAGATCAAAAGATCAGCTACCGCATCTTGACA 1510
Qy 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1511 GACCCAGCAGGTGGTGTAGCCATGGACCCAGACAGTGGGACAGTCCAGCTGTGGGAC 1570
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1571 CTGACCGTGGAGTGAAGCAGTTTGTGAGGAACACATCTATGAAGTCACTGTCTTGGCC 1630
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db 1631 ATGCACATGGAAGCCCTCCACACTGCACGGGAAACCTTCTGTCTAACACTGATTGAT 1690
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1691 GTCAATGACCATGGCCCGAGTCCCTGAGCCCCGTGAGATCACCATCTGCAACCAAGCCCT 1750
Qy 561 ValArgHis-ValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1751 GTGGGCCA-GGTGCTGAACATCAGCGACAAGGACCTGTCTCCCAACACCTCCCTTCCA 1809
Qy 580 nAlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGluGlyAs 600
Db 1810 GGCCAGCTCACAGATGACTCAGACATCTACTGACGCGAGAGGTCAACGAGGAGGTGA 1869
Qy 600 pThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSe 620
Db 1870 CACAGTGGTCTTGTCTCCCTGAGAGTTCTTGAAGCAGGATACATATGACGTGCACCTTTC 1929
Qy 620 rLeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCy 640
Db 1930 TCTGTCTGACCATGGCAACAAGAGCAGCTGACGTGATCAGGGCCACTGTGTGGAGTGA 1989
Qy 640 sHisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLe 660
Db 1990 CCATGGCCATGTCAAAACCTGCTCCCTGGACCTTGGAAAGGAGGTTCATCTCTCCCTGTGT 2049
Qy 660 uGlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValValArgLysLys 680
Db 2050 GGGGGCTGCTGGCTCTGCT 2109
Qy 680 sArgLysIleLysGluProLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 700
Db 2110 GCGGAAGATCAAGAGGCCCT 2169
Qy 700 rTyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisAr 720
Db 2170 CTATGGCGAAGAGGGGGTGGCGAAGAGGACGAGGACTATGACATCACCCAGCTCCACCG 2229
Qy 720 sGlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLepr 740
Db 2230 AGGTCTGGAGGCCAGCGCGAGGTGTTCTTCGCAATGACGTGGGACCAACCATCATCCC 2289
Qy 740 oThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGl 760

CC work as modulators of Ras activity by inducing expression of tumour
 CC suppressor genes. (1), and polypeptides encoded by them, are useful as
 CC targets for diagnosis or therapy and in screening to determine the
 CC effects of an active compound (potential pharmaceutical) on a cell line,
 CC particularly for diagnosis and treatment of tumors, especially by
 CC modulating expression of (1) (by gene therapy, antisense RNA or ribozyme
 CC methods) or by modulating the amount and/or location of (1)-encoded
 CC polypeptides (by administration of the polypeptide or its activator,
 CC antibody (optionally as a conjugate) or inhibitor). The method allows
 CC identification of many class II tumour suppressor genes (i.e. genes that
 CC are not primary targets for tumour-initiating mutations). AAH81492-
 CC AAH82376 represent the human and rat derived nucleic acid fragments
 CC described in the method of the invention
 XX

SQ Sequence 306 BP; 73 A; 110 C; 73 G; 50 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.55e-86	Length:	306
Score:	102.00	Matches:	102
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	12.30%	Indels:	0
DB:	5	Gaps:	0

US-09-916-849A-1 (1-829) x AAH81822 (1-306)

QY	707	GlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArgGlyLeuGluAlaArgPro	726
Db	1	GGCGAAGAGGACGAGACTATGACATCACCCAGCTCCACCGAGGTCTGGAGGCCAGGCCG	60
QY	727	GluValValLeuArgAsnAspValAlaProThrIleIleProThrProMetTyrArgPro	746
Db	61	GAGGTGGTTCCTCCGCAATGACGTGGCACCACCATCCCGACACCCATGTACCGTCCT	120
QY	747	ArgProAlaAsnProAspGluIleGlyAsnPheIleIleGluAsnLeuIleAlaAlaAsn	766
Db	121	CGGCCAGGCCAACCCAGATGAATCGCAACTTTATTAATGAGAACCTGAGGCGGCTAAC	180
QY	767	ThrAspProThrAlaProProTyrAspThrLeuLeuValPheAspTyrGluGlySerGly	786
Db	181	ACAGACCCACAGCCCGCCCTACGACACCCCTCTTGGTGTTCGACTATGAGGGCAGCGGC	240
QY	787	SerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSerAspGlnAspTyr	806
Db	241	TCCGACGCGCGCTCCCTGAGCTCCCTCACCTCCTCCGCTCCGACCAAGACCAAGATTAC	300
QY	807	AspTyr	808
Db	301	GATTAT	306

Search completed: September 23, 2004, 03:59:03
 Job time : 855 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 23, 2004, 03:29:35 ; Search time 144 Seconds
(without alignments)
3194.821 Million cell updates/sec

Title: US-09-916-849a-1

Perfect score: 829

Sequence: 1 MGLPRGLASLLQLVCWLQ.....NEWGSRFKKLMYGGGDD 829

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1360453

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters: -DEV=xlh

-O=/cgn2_1/USPTO/spool_h/US09916849/runat_21092004_145836_7341/app_query.fasta_1.967
-DB=Issued_Patents_NA -QFWT=fastap -SUFFIX=oligo.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdd
-LIST=45 -DOALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09916849 @CNC 1 1.69 @runat_21092004_145836_7341 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=7 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	829	100.0	3171	4	US-09-016-434-1348
2	14	1.7	2808	1	US-08-237-919-1
3	14	1.7	2808	4	US-08-732-429-1
4	14	1.7	2808	4	US-09-798-267-1
5	14	1.7	2808	5	PCT-US95-05518-1
6	13	1.6	3048	1	US-08-188-228-47
7	13	1.6	3048	1	US-08-332-643-41
8	13	1.6	3048	1	US-08-332-638-47
9	12	1.4	436	1	US-09-397-787-318
10	11	1.3	117	1	US-08-188-228-9
11	11	1.3	117	1	US-08-332-643-9
12	11	1.3	117	1	US-08-332-638-9

13	11	1.3	938	4	US-09-976-594-1052	Sequence 1052, Ap
14	11	1.3	1369	1	US-08-188-228-51	Sequence 51, Appl
15	11	1.3	1369	1	US-08-332-643-45	Sequence 45, Appl
16	11	1.3	1369	1	US-08-332-638-51	Sequence 51, Appl
17	10	1.2	326	4	US-09-389-681-288	Sequence 288, App
18	10	1.2	326	4	US-09-620-4058-288	Sequence 288, App
19	10	1.2	326	4	US-09-339-338-288	Sequence 288, App
20	10	1.2	326	4	US-09-433-8268-288	Sequence 288, App
21	10	1.2	326	4	US-09-604-287A-288	Sequence 288, App
22	10	1.2	326	4	US-09-834-759-288	Sequence 288, App
23	9	1.1	402	1	US-08-346-611-1	Sequence 1, Appli
24	9	1.1	402	1	US-08-346-611-3	Sequence 3, Appli
25	9	1.1	402	1	US-08-794-494-1	Sequence 1, Appli
26	9	1.1	402	2	US-08-794-494-3	Sequence 3, Appli
27	9	1.1	813	4	US-09-641-638-100	Sequence 100, App
28	9	1.1	943	4	US-09-641-638-99	Sequence 99, Appl
29	9	1.1	1484	2	US-08-037-816A-17	Sequence 17, Appl
30	9	1.1	1484	2	US-08-037-816A-21	Sequence 21, Appl
31	9	1.1	1484	2	US-08-530-146-17	Sequence 17, Appl
32	9	1.1	1484	2	US-08-530-146-21	Sequence 21, Appl
33	9	1.1	1571	2	US-08-037-816A-25	Sequence 25, Appl
34	9	1.1	1571	2	US-08-530-146-25	Sequence 25, Appl
35	9	1.1	1818	4	US-09-475-515-59	Sequence 59, Appl
36	9	1.1	1818	4	US-09-475-515-61	Sequence 61, Appl
37	9	1.1	1818	4	US-09-475-515-62	Sequence 62, Appl
38	9	1.1	1836	4	US-09-475-515-41	Sequence 41, Appl
39	9	1.1	1836	4	US-09-475-515-44	Sequence 44, Appl
40	9	1.1	1836	4	US-09-475-515-47	Sequence 47, Appl
41	9	1.1	1863	4	US-09-475-515-63	Sequence 63, Appl
42	9	1.1	1944	4	US-09-475-515-37	Sequence 37, Appl
43	9	1.1	1944	4	US-09-475-515-38	Sequence 38, Appl
44	9	1.1	1944	4	US-09-475-515-40	Sequence 40, Appl
45	9	1.1	1944	4	US-09-475-515-43	Sequence 43, Appl

ALIGNMENTS

RESULT 1
US-09-016-434-1348
; Sequence 1348, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: FA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555

TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1348:
SEQUENCE CHARACTERISTICS:
LENGTH: 3171 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: G35322
US-09-016-434-1348

Alignment Scores:		
Pred. No.:	0	Length: 3171
Score:	829.00	Matches: 829
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	100.00%	Indels: 0
DB:	4	Gaps: 0

US-09-916-849A-1 (1-829) x US-09-016-434-1348 (1-3171)

1	MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln	20
54	ATGGGGCTCCCTCGTGGACCTCTGGCTCTCTCTCTCTCAGGTTTGTGGCTGCAG	113
21	CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla	40
114	TGCGGGGCTCCGACCGTGGCGGGGTCTTCAAGGAGGCTGAGTGACCTTGGAGGGC	173
41	GlyGlyAlaGluGlnGluProGlyClnAlaLeuGlyLysValPheMetGlyCysProGly	60
174	GGAGCGCGGAGCAGGACCGCGCGCAGCGCTGGGGAAGTATTTCATGGGCTCCCTCGG	233
61	GlnGluProAlaLeuPheSerThrAsnAsnAspPheThrValArgAsnGlyGluThr	80
234	CAAGAGCCACGCTCTGTTTAGCACTGATATGATGACTTCACTGTGGGAATGGCGAGCA	293
81	ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg	100
294	GTCCAGGAAAGAGGTCACTCAAGGAAGGAATCCATTGAAGATCTTCCCATCCAAAGT	353
101	IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluLysGly	120
354	ATCTTACGAAGACACAGAGAGATTGGGTGGTCTCCAAATATCTGTCCCTGAAAATGGC	413
121	LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys	140
414	AAGGGTCCTTCCCCACAGACTCAATCAGCTCAAGTCTAATAAGATAGACACCAAG	473
141	IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal	160
474	ATTTTCTACAGCATCACGGGCGGGGCGACAGCCCCCTGAGGTTCTTCGTGTGTA	533
161	GluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys	180
534	GAGAAGGAGACAGGCTGTTGTTGTAATGAAGCACTGGACCGGGAGAGATTGCCAAG	593
181	TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn	200
594	TATGAGCTCTTGGCCAGCTGTGTACAGAAATGGTGCCTCAGTGGAGGACCCCAATGAA	653
201	IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe	220
654	ATCTCCATCATCGTACCGCCAGAAATGACCACAAGCCCAAGTTTACCAGGACACCTTC	713
221	ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr	240
714	CGAGGGAGTGCTTAGGGAGTCCATCAGAGTACTTCTGTGATGAGGTGACGAGCCACA	773
241	AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln	260
774	GATGAGGATGATGCCATCTCACTACATGGGGTGGTGTCTTACTTCCATATAGCCAA	833

QY 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
DB 1914 CTGTCTGACCATGGCAACAAGAGCAGCTGACGGTGATCAGGGCCACTGTGTGGACTGC 1973
QY 641 HisGlyHisValGluThrCysProGlyProThrLysGlyGlyPheIleLeuProValLeu 660
DB 1974 CATGCCCATGTGGAACCTGCCCTGGACCTGGAAGAGGTTTCATCTCCCTGTGTG 2033
QY 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuValLeuValArgLysLys 680
DB 2034 GGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2093
QY 681 ArgLysIleLysGluProLeuLeuLeuProGluAspThrArgAsnValPheTyr 700
DB 2094 CGGAAGATCAAGAGGCCCTCTACTCCAGAGATGACCCCGTGACAACGCTCTTCTAC 2153
QY 701 TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
DB 2154 TATGGGAGAGGGGGTGGCGAGAGGACGAGGACTATGATCACCAGCTCCACCGA 2213
QY 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
DB 2214 GGTCTGGAGGCCAGCGCGAGGTGTTCTCCGAATGACGTGGCACCACCAACCATCCTCCG 2273
QY 741 ThrProMetTyrArgProArgProAlaAsnProAlaAsnProGluIleGlyAsnPheIleGlu 760
DB 2274 ACACCATGTACCGTCTAGGCAGCCACCCAGATGAATCGGCACTTTATTAATTGAG 2333
QY 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
DB 2334 AACCTGAAGGGGCTAAACACAGACCCACAGCCCGCCCTACGACACCTCTTGTGTGTC 2393
QY 781 AspTyrGluGlySerGlySerAspAlaAsnLeuSerSerLeuThrSerSerAlaSer 800
DB 2394 GACTATGAGGAGGCGGTCCGACCGCGTCCCTGAGCTCCCTCACCCTCTCCGCTCC 2453
QY 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
DB 2454 GACCAAGACCAAGATACGATTATCTGACGAGTGGGCGAGCGGTTCAAGAGCTGCA 2513
QY 821 AspMetTyrGlyGlyGlyGluAspAsp 829
DB 2514 GACATGTACGTGGTGGGAGGAGCAGC 2540

RESULT 2
US-08-237-919-1
; Sequence 1, Application US/08237919
; Patent No. 5610281
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael B
; TITLE OF INVENTION: Methods and Compositions for
; Modulating Heterotypic E-cadherin Interactions with T Lymphocytes
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/237,919
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R
; REGISTRATION NUMBER: 36,637

REFERENCE/DOCKET NUMBER: B0801/7023
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2808 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: liver
FEATURE:
NAME/KEY: CDS
LOCATION: 109..2745
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 109..558
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 559..2742
FEATURE:
NAME/KEY: cytoplasmic tail
LOCATION: 2296..2746
FEATURE:
NAME/KEY: HAV tripeptide
LOCATION: 800..808
FEATURE:
NAME/KEY: transmembrane
LOCATION: 2225..2295
US-08-237-919-1
Alignment Scores:
Pred. No.: 0.0011 Length: 2808
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.69% Indels: 0
DB: 1 Gaps: 0
US-09-916-849A-1 (1-829) x US-08-237-919-1 (1-2808)
QY 815 ArgPheLysLysLeuAlaAspMetTyrGlyGlyGluAsp 828
DB 2698 CGCTTCAAGAGCTGGCTGACATGTACGGAGCGCGGAGGAC 2739
RESULT 3
US-08-732-429-1
; Sequence 1, Application US/08732429
; Patent No. 6300080
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic
; E-cadherin Interactions with T Lymphocytes
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/732,429

```
; FILING DATE: herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/237,919
; FILING DATE: 3 May 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2808 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: liver
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 109..2745
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 109..558
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 559..2742
; FEATURE:
; NAME/KEY: cytoplasmic tail
; LOCATION: 2296..2746
; FEATURE:
; NAME/KEY: HAV tripeptide
; LOCATION: 800..808
; FEATURE:
; NAME/KEY: transmembrane
; LOCATION: 2225..2295
; US-08-732-429-1
Alignment Scores:
Pred. No.: 0.0011 Length: 2808
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.69% Indels: 0
DB: 4 Gaps: 0
US-09-916-849A-1 (1-829) x US-08-732-429-1 (1-2808)
QY 815 ArgpHelvSLysLeuAlaAspMetTyvRGlyGlyGluasp 828
Db 2698 CGCTTCAGAGACTGGCTGACATGTACGAGGCGCGGAGGAC 2739
RESULT 4
US-09-798-267-1
; Sequence 1, Application US/09798267
; Patent No. 6406870
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael
; APPLICANT: Ceppek, Karyn
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic E-cadherin
; FILE REFERENCE: L0560/7008ERP
; CURRENT APPLICATION NUMBER: US/09798,267
; CURRENT FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 08/237,919
; PRIOR FILING DATE: 1994-05-03
; PRIOR APPLICATION NUMBER: PCT/US 95/05518
; PRIOR FILING DATE: 1995-05-03
```

```
; PRIOR APPLICATION NUMBER: US 08/732,429
; PRIOR FILING DATE: 1996-11-01
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2808
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(2742)
; NAME/KEY: mat_peptide
; LOCATION: (2740)..()
; NAME/KEY: misc_feature
; LOCATION: (800)..(808)
; OTHER INFORMATION: HAV tripeptide
; NAME/KEY: misc_feature
; LOCATION: (2225)..(2295)
; OTHER INFORMATION: transmembrane
; NAME/KEY: misc_structure
; LOCATION: (2296)..(2746)
; OTHER INFORMATION: cytoplasmic tail
; NAME/KEY: sig_peptide
; LOCATION: (109)..(558)
; US-09-798-267-1
Alignment Scores:
Pred. No.: 0.0011 Length: 2808
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.69% Indels: 0
DB: 4 Gaps: 0
US-09-916-849A-1 (1-829) x US-09-798-267-1 (1-2808)
QY 815 ArgpHelvSLysLeuAlaAspMetTyvRGlyGlyGluasp 828
Db 2698 CGCTTCAGAGACTGGCTGACATGTACGAGGCGCGGAGGAC 2739
RESULT 5
PCT-US95-05518-1
; Sequence 1, Application PC/TUS9505518
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Methods and Compositions for Modulating Heterotypic
; E-cadherin Interactions with T Lymphocytes
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05518
; FILING DATE: herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/237,919
; FILING DATE: 3 May 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
```

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2808 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: liver
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 109..2745
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 109..558
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 559..2742
; NAME/KEY: cytoplasmic tail
; LOCATION: 2296..2746
; FEATURE:
; NAME/KEY: HAV tripeptide
; LOCATION: 800..808
; FEATURE:
; NAME/KEY: transmembrane
; LOCATION: 2225..2295
; PCT-US95-05518-1
;
; Alignment Scores:
; Pred. No.: 0.0011 Length: 2808
; Score: 14.00 Matches: 14
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 1.69% Indels: 0
; DB: 5 Gaps: 0
;
; US-09-916-849A-1 (1-829) x PCT-US95-05518-1 (1-2808)
;
; QY 815 ArgPheLysLeuAlaAspMetTyrGlyGlyGluAsp 828
; DB 2698 CGCTTCAAGAGCTGGCTGACATGTACGGAGCGCGAGGAC 2739
;
; RESULT 6
; US-08-188-228-47
; Sequence 47, Application US/08188228
; Patent No. 5597725
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,228
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/049,460
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3048 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3048 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-188-228-47
;
; Alignment Scores:
; Pred. No.: 0.0113 Length: 3048
; Score: 13.00 Matches: 13
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 1.57% Indels: 0
; DB: 1 Gaps: 0
;
; US-09-916-849A-1 (1-829) x US-08-188-228-47 (1-3048)
;
; QY 815 ArgPheLysLeuAlaAspMetTyrGlyGlyGlu 827
; DB 2718 AGATTCAGAGAGCTGGGACATGTATGGAGGTGTGAA 2756
;
; RESULT 7
; US-08-332-643-41
; Sequence 41, Application US/08332643
; Patent No. 5639634
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; Bicknell
; STREET: Two First National Plaza, 20 South Clark
; STREET: Street
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,643
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/872,643
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5639634ard, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 41:
; APPLICATION NUMBER: US 08/049,460

```

```

/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3048 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
US-08-332-643-41
Alignment Scores:
Pred. No.: 0.0113 Length: 3048
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.57% Indels: 0
DB: 1 Gaps: 0
US-09-916-849A-1 (1-829) x US-08-332-643-41 (1-3048)
RESULT 8
QY 815 ArgPheLysLeuAlaAspMetTyrGlyGlyGlu 827
DB 2718 AGATTCAAGAAGCTGGCGACATGTATGGAGGTGTGAA 2756
SEQUENCE 47, Application US/08332638
Patent No. 5646250
GENERAL INFORMATION:
APPLICANT: Suzuki, Shintaro
TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Borun
STREET: 6300 Sears Tower, 233 S. Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/332,638
FILING DATE: 01-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,643
FILING DATE: 17 APR 1992
APPLICATION NUMBER: US/08/049,460
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: No. 5646250and, Greta E.
REGISTRATION NUMBER: 35,502
REFERENCE/DOCKET NUMBER: 31340
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 3048 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-332-638-47
Alignment Scores:
Pred. No.: 0.0113 Length: 3048
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0

```

```

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.57% Indels: 0
DB: 1 Gaps: 0
US-09-916-849A-1 (1-829) x US-08-332-638-47 (1-3048)
QY 815 ArgPheLysLeuAlaAspMetTyrGlyGlyGlu 827
DB 2718 AGATTCAAGAAGCTGGCGACATGTATGGAGGTGTGAA 2756
RESULT 9
US-09-397-787-318
SEQUENCE 318, Application US/09397787
Patent No. 6468758
GENERAL INFORMATION:
APPLICANT: Benson, Darin R.
APPLICANT: Lodes, Michael J.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: King, Gordon E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
FILE REFERENCE: 210121.466C2
CURRENT FILING DATE: 1999-09-16
CURRENT APPLICATION NUMBER: US/09/397,787
NUMBER OF SEQ ID NOS: 334
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 318
LENGTH: 436
TYPE: DNA
ORGANISM: Homo sapien
US-09-397-787-318
Alignment Scores:
Pred. No.: 0.0162 Length: 436
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.45% Indels: 0
DB: 4 Gaps: 0
US-09-916-849A-1 (1-829) x US-09-397-787-318 (1-436)
QY 815 ArgPheLysLeuAlaAspMetTyrGlyGlyGly 826
DB 290 CGGTTCAAGAACTTGTGACATGTATGGAGGT 325
RESULT 10
US-08-188-228-9
SEQUENCE 9, Application US/08188228
Patent No. 5597725
GENERAL INFORMATION:
APPLICANT: Suzuki, Shintaro
TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
ADDRESSEE: Borun
STREET: 6300 Sears Tower, 233 S. Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/188,228
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/049,460

```

```

; FILING DATE: 19 APR 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,643
; FILING DATE: 17 APR 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: NO. 5597725and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31340
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-188-228-9

Alignment Scores:
Pred. No.: 0.0427 Length: 117
Score: 11.00 Matches: 11
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.33%
Indels: 0
DB: 1
Gaps: 0

US-09-916-849A-1 (1-829) x US-08-188-228-9 (1-117)

QY 777 LeuLeuValPheAspTyrGluGlySerGlySer 787
Db 4 CTGCTGGTCTTCGACTACGAAGCAGCGGTCTT 36

RESULT 11
US-08-332-643-9
; Sequence 9, Application US/08332643
; Patent No. 5639634
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,643
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/872,643
; FILING DATE: 17 APR 1992
; APPLICATION NUMBER: US/08/049,460
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5639634and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 984-9740
; TELEFAX: 25-3856
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-332-638-9

Alignment Scores:
Pred. No.: 0.0427 Length: 117
Score: 11.00 Matches: 11
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.33%
Indels: 0
DB: 1
Gaps: 0

US-09-916-849A-1 (1-829) x US-08-332-643-9 (1-117)

QY 777 LeuLeuValPheAspTyrGluGlySerGlySer 787
Db 4 CTGCTGGTCTTCGACTACGAAGCAGCGGTCTT 36

RESULT 12
US-08-332-638-9
; Sequence 9, Application US/08332638
; Patent No. 5646250
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,638
; FILING DATE: 01-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,643
; FILING DATE: 17 APR 1992
; APPLICATION NUMBER: US/08/049,460
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5646250and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31340
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-332-638-9

Alignment Scores:
Pred. No.: 0.0427 Length: 117
Score: 11.00 Matches: 11
Percent Similarity: 100.00%
```

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.33% Indels: 0
DB: 1 Gaps: 0

US-09-916-849A-1 (1-829) x US-08-332-638-9 (1-117)

QY 777 LeuLeuValPheAspTyrGluGlySerGlySer 787
|||||
DB 4 CTGCTGCTTCGACTACGAGCGCGGTCT 36

RESULT 13

US-09-976-594-1052
; Sequence 1052, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 1052
; LENGTH: 938
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 272273.1
US-09-976-594-1052

Alignment Scores: 938
Pred. No.: 0.321 Length: 938
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.33% Indels: 0
DB: 4 Gaps: 0

US-09-916-849A-1 (1-829) x US-09-976-594-1052 (1-938)

QY 815 ArgPheLysLeuAlaAspMetTyrGlyGly 825
|||||
DB 349 CGTTTCAAAAAGCTTCCGATATGATGGGGT 381

RESULT 14

US-08-188-228-51
; Sequence 51, Application US/08188228
; Patent No. 5597725
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,228
; FILING DATE:
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/049,460
; FILING DATE: 19 APR 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,643
; FILING DATE: 17 APR 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5597725and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31340
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1369 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-188-228-51
Alignment Scores:
Pred. No.: 0.462 Length: 1369
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.33% Indels: 0
DB: 1 Gaps: 0

US-09-916-849A-1 (1-829) x US-08-188-228-51 (1-1369)

QY 815 ArgPheLysLeuAlaAspMetTyrGlyGly 825
|||||
DB 1190 CGATTCAAAAAGCTTGCAGATATGATGGAGGA 1222

RESULT 15

US-08-332-643-45
; Sequence 45, Application US/08332643
; Patent No. 5639634
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza, 20 South Clark
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,643
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/872,643
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5639634and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/30795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 346-5750
; TELEFAX: (312) 984-9740


```

; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1369 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
US-08-332-643-45

Alignment Scores:
  Pred. No.:      0.462      Length:      1369
  Score:          11.00      Matches:      11
  Percent Similarity: 100.00%  Conservative: 0
  Best Local Similarity: 100.00%  Mismatches: 0
  Query Match:      1.33%      Indels:      0
  DB:                1        Gaps:        0

US-09-916-849A-1 (1-829) x US-08-332-643-45 (1-1369)

QY      815  ArgPheLysLysLeuAlaAspMetTyGlyGly 825
      |||||
Db      1190  CGATTCAAAAGCTTCAGATATGATGAGGA 1222

```

Search completed: September 23, 2004, 07:41:11
 Job time : 176 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: September 23, 2004, 06:21:13 ; Search time 1807 Seconds
(without alignments)
2323.654 Million cell updates/sec

Title: US-09-916-849A-1

Perfect score: 829

Sequence: 1 MGLPRGPLASILLQLVQCWLQ.....NEWGSRFKLADMYGGEDD 829

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3337386 seqs, 2532474682 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6665927

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool.h/US0916849/runat_21092004_145838_7417/app.query.fasta_1.967
-DS=Published Applications NA -QMT=fastcap -SUFFIX=oligo.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -SCAP=1 -END=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFM=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US0916849 @CGN 1.1.593 @runat_21092004_145838_7417
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/prodata/2/pubpna/PT_NEW_PUB.seq:
3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/prodata/2/pubpna/PTUS_PUBCOMB.seq:
7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:
14: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
*					

1	829	100.0	3170	10	US-09-975-723A-2	Sequence 2, Appli
2	829	100.0	3171	9	US-09-905-983-49	Sequence 49, Appl
3	829	100.0	3171	9	US-09-880-107-3759	Sequence 3759, Ap
4	829	100.0	3171	13	US-10-058-270A-125	Sequence 125, App
5	829	100.0	3171	13	US-10-342-887-669	Sequence 669, App
6	829	100.0	3171	13	US-10-172-118-669	Sequence 669, App
7	829	100.0	3171	15	US-10-174-677-115	Sequence 115, App
8	829	100.0	3171	15	US-10-158-123-2	Sequence 2, Appli
9	829	100.0	3171	16	US-10-295-027-782	Sequence 782, App
10	829	100.0	3171	16	US-10-295-027-799	Sequence 799, App
11	829	100.0	3171	16	US-10-295-027-862	Sequence 862, App
12	829	100.0	3171	16	US-10-295-027-895	Sequence 895, App
13	829	100.0	3171	16	US-10-305-720-1348	Sequence 1348, Ap
14	829	100.0	3171	17	US-10-775-920-188	Sequence 188, App
15	829	100.0	3777	13	US-10-116-802-6	Sequence 6, Appli
16	728	87.8	3205	13	US-10-229-345-7	Sequence 7, Appli
17	728	87.8	3205	13	US-10-274-177-7	Sequence 7, Appli
18	728	87.8	3205	15	US-10-301-822-17	Sequence 17, Appl
19	728	87.8	3205	16	US-10-295-027-337	Sequence 337, App
20	728	87.8	3205	16	US-10-295-027-1029	Sequence 1029, Ap
21	728	87.8	3205	16	US-10-087-080-13	Sequence 13, Appl
22	728	87.8	3205	17	US-10-650-112-7	Sequence 7, Appli
23	728	87.8	3205	17	US-10-775-920-186	Sequence 186, App
24	715	86.2	3219	9	US-09-895-652-17	Sequence 17, Appl
25	659	79.5	3632	17	US-10-775-920-187	Sequence 187, App
26	102	12.3	306	10	US-09-930-213-331	Sequence 331, App
27	89	10.7	3314	16	US-10-062-674-1863	Sequence 1863, Ap
28	80	9.7	247	15	US-10-029-386-15020	Sequence 15020, A
29	80	9.7	555	15	US-10-029-386-1318	Sequence 1318, Ap
30	69	8.3	210	15	US-10-029-386-17424	Sequence 17424, A
31	69	8.3	574	15	US-10-029-386-3724	Sequence 3724, Ap
32	54	6.5	255	10	US-09-930-213-381	Sequence 381, App
33	52	6.3	160	15	US-10-158-123-3	Sequence 3, Appli
34	47	5.7	406	16	US-10-027-632-571	Sequence 571, App
35	30	3.6	767	13	US-10-062-674-16189	Sequence 16189, A
36	30	3.6	767	16	US-10-027-632-16189	Sequence 16189, A
37	21	2.5	522	9	US-09-920-455-77	Sequence 77, Appl
38	19	2.3	60	10	US-09-908-975-13599	Sequence 13599, A
39	15	1.8	195	15	US-10-029-386-16396	Sequence 16396, A
40	15	1.8	542	15	US-10-029-386-2696	Sequence 2696, Ap
41	15	1.8	1351	15	US-10-198-846-12909	Sequence 12909, A
42	15	1.8	2768	9	US-09-905-983-4	Sequence 4, Appli
43	15	1.8	2768	9	US-09-905-983-6	Sequence 6, Appli
44	15	1.8	4396	12	US-10-152-319A-1900	Sequence 1900, Ap
45	15	1.8	4778	13	US-10-342-887-1027	Sequence 1027, Ap

ALIGNMENTS

RESULT 1
US-09-975-723A-2
; Sequence 2, Application US/09975723A
; Publication No. US20030108529A1
; GENERAL INFORMATION:
; APPLICANT: Nackman, Gary
; APPLICANT: Foly, Ramsey
; TITLE OF INVENTION: Improvement of Endothelial Cell-Cell
; TITLE OF INVENTION: Conesion
; FILE REFERENCE: 601-1-101N
; CURRENT APPLICATION NUMBER: US/09/975,723A
; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 60/241,216
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/243,693
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 3170
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-975-723A-2

Alignment Scores:					
Pred. No.:	0	Length:	3170		
Score:	829.00	Matches:	829		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	100.00%	Indels:	0		
DB:	10	Gaps:	0		
US-09-916-849A-1 (1-829) x US-09-975-723A-2 (1-3170)					
Qy	1	MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuValCysTrpLeuGln	20		
Ds	54	ATGGGCTCCCTCGTGGACCTCTCGCGTCCTCTCTCTCTCAGGTGTGCTGGCTGCAG	113		
Qy	21	CyeAlaAsaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluLa	40		
Ds	114	TGGCGGGCTCCGAGCGGTCGGGCGGTCTTCAGGGAGGCTGAAGTGACTTTGGAGGG	173		
Qy	41	GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly	60		
Ds	174	GGAGGCGGAGCAGGAGCCCGGCCGCTGGGAAAGTAATTATGGCTGCCCTGGG	233		
Qy	61	GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr	80		
Ds	234	CAAGAGCCAGCTCTGTTAGCACTCATTAATGATGACTTCTGTGGGAATGGCGAGA	293		
Qy	81	ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg	100		
Ds	294	GTCAGGAAGAAGGTCACTGAGGARAGGAATCCATTGAAGATCTTCCATCCAACGT	353		
Qy	101	IleLeuArgHisLysArgAspTrpValValAlaProLeuSerValProGluAsnGly	120		
Ds	354	ATCTTAGAAGACACAAGAGAGATTGGGTGGTCTCTCCAATATCTGTCCCTGAAAATGGC	413		
Qy	121	LysGlyProPheProGluArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys	140		
Ds	414	AAGGTCCTCTCCCCAGAGACTGAATCAGCTCAAGTCTAATAAGATAGAGACCACAG	473		
Qy	141	IlePheTyrrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal	160		
Ds	474	ATTTCCTACAGCATCACGGGGCGGGGCGACAGCCCCCTCAGGGTGTCCTCGCTGTA	533		
Qy	161	GluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluLeuAlaLys	180		
Ds	534	GAGAAGAGACAGGCTGGTGTGTGTGAATAGCCACTGGACCGGAGNGAGATTGCCAG	593		
Qy	181	TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn	200		
Ds	594	TATGAGCTCTTTGGCCAGCGCTGTGTCAGAGAATGGTCCTCAGTGGAGGACCCCATGAAC	653		
Qy	201	IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe	220		
Ds	654	ATCTCCATCATGTGACCGACCGAATGACCACAGCCCAAGTTTACCCAGGACCTTC	713		
Qy	221	ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr	240		
Ds	714	CGAGGGAGTGTCTTAGGGGAGTCTACAGGTACTTCTGTGATGAGGTGACAGACACA	773		
Qy	241	AspGluAspAlaIleIleThrTyrrAsnGlyValValAlaTyrrSerIleHisSerGln	260		
Ds	774	GATGAGGATGATGCCATCTACACCTACAAATGGGGTGGTGTCTTACTTCATCCATAGCAA	833		
Qy	261	GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer	280		
Ds	834	GAACCAAGGACCCACAGACCTCATGTCTCAATTCACCGGAGCACAGGACCATCAGC	893		
Qy	281	ValIleSerSerGlyLeuAspArgGluLysValProGluTyrrThrLeuThrIleGlnAla	300		
Ds	894	GTCAATCTCAGTGGCTGGACCGGAAAAAGTCCCTGAGTACACACTGACCATCCAGGCC	953		
Qy	301	ThrAspMetAspGlyAspGlySerThrThrThraValAlaValGluIleLeuLeuAsp	320		

```
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspThrArgAspAsnValPheTyr 700
Db 2094 CGAAGATCAAGGAGCCCTCTTACTCCCAAGAAGATGACACCGGTGACAAGGTCTTCTAC 2153
Qy 701 TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnIleuHisArg 720
Db 2154 TATGGCGAAGAGGGGGTGGCAAGAGGACCAAGGATGACATCACCGAGTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db 2214 GGTCTGGAGGCCAGCGGAGGTGGTCTCCGCAATGACGTGGCACCACCATCATCCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPhelleIleGlu 760
Db 2274 ACACCATGTACCGTCTCTAGCCGAGCCACCAAGATGAATCGCAACTTTTAAATTGAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGNAGCGGCTAACACAGACCCCAAGCCCGCCCTACGACACCTCTTGGTGTTC 2393
Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGCGAGCGCTCCGACGCGCTCCCTGAGCTCCCTCACCTCCCTCCGCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGACAGTGGGGAGCGCTTCAAGAGCTGGCA 2513
Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 2514 GACATGTACGGTGGCGGGAGGACGAC 2540

RESULT 2
US-09-905-983-49
; Sequence 49, Application US/09905983
; Patent No. US20020045591A1
; GENERAL INFORMATION:
; APPLICANT: Geiger, Benjamin
; APPLICANT: Ben-Ze'ev, Avri
; APPLICANT: Sadot, Einat
; TITLE OF INVENTION: METHODS AND THERAPEUTIC COMPOSITIONS FOR TREATING CANCER
; FILE REFERENCE: 01/22326
; CURRENT APPLICATION NUMBER: US/09/905,983
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49
; LENGTH: 3171
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-905-983-49

Alignment Scores:
Pred. No.: 0 Length: 3171
Score: 829.00 Matches: 829
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-916-849a-1 (1-829) x US-09-905-983-49 (1-3171)
Qy 1 MetGlyLeuProArgGlyProLeuAlaSerIleuLeuLeuGlnValCysTrpLeuGln 20
Db 54 ATGGGGCTCCCTCGTGAGCCTCTCGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113
Qy 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
Db 114 TGGCGGCGCTCCGAGCGGTCCGGGGGCTCTTCAGGAGGCTTGAAGTGAACCTTGGAGGG 173
Qy 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db 115 TGGCGGCGCTCCGAGCGGTCCGGGGGCTCTTCAGGAGGCTTGAAGTGAACCTTGGAGGG 173
```

```
Db 174 GGAGCGCGGAGCAGGAGCCCGCCAGCGGTGGGGAAAGATTATTCATGGGTGCCTGGG 233
Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspAspPheThrValArgAsnGlyGluThr 80
Db 234 CAAGAGCCAGCTCTGTTTAGCAGCTGATATGATGATGATGATGATGATGATGATGATG 293
Qy 81 ValGlnGluArgArgSerIleuLysGluArgAsnProLeuLysIlePheProSerIysArg 100
Db 294 GTCCAGGAAGAAGGTCTCACTGAAGGAAGAATCCATTTGAAGATCTTCCCATCCAAAGC 353
Qy 101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
Db 354 ATCTTACGACACACACAGAGATTTGGGTGGTTCCTCAATATCTGTCCTGAAATGGC 413
Qy 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrIys 140
Db 414 AAGGTCTCTCTCCCGACAGACTGAATCAGCTCAAGTCTAATAAAGATAGACACCAAG 473
Qy 141 IlePheTyrSerIleThrGlyProGlyValaAspSerProGluGlyValPheAlaVal 160
Db 474 ATTTTCTACAGCATCACGGGGCCGGGGCAGACAGCCCCCTGAGGGTGTCTTCTGCTGA 533
Qy 161 GluLysGluThrGlyTyrIleLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
Db 534 GAGAAGGAGACAGGCTGTTGTTGAATTAAGCCACTGGACCGGAGGAGATTGCCAAG 593
Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 594 TATGAGCTCTTTGGCCAGCTGTGTACAGAAATGGTGTCTCAGTGGAGGACCCCATGAAC 653
Qy 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 654 ATCTCATCATCGTGACCGACCAAGATGACCAAGCCCAAGTTTACCAGGACACCTTC 713
Qy 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 714 CGAGGAGTGTCTTAGAGGAGTCTCTACAGGTACTTCTGTGATGATGATGATGATGATGAT 773
Qy 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
Db 774 GATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 833
Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 834 GAAACCAAGGACCCACACGCTCATGTTTCAATTCACCGGAGCAGCAGCAGCAGCAGCAG 893
Qy 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db 894 GTCATCTCCAGTGGCCTGGACCGGAAAGTCCCTGAGTACACACATCACCATCCAGGCC 953
Qy 301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValValGluIleLeuAsp 320
Db 954 ACAGACATGGATGGGACGGCTCCACACACCGGAGTGGCAGTAGTGGAGATCCTTGAT 1013
Qy 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1014 GCCAATGACAAATGCTCCCATGTTTACCCCGCAGAGTACGAGGCCCATGTGCTCCGAGA 1073
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1074 CCAGTGGGCCCATGAGGTGCAGAGCTGACGCTCACTGNTCTGGACGCCCCCACTCACA 1133
Qy 361 AlaTrpArgAlaThrTyrLeuIleValGlyAspAspGlyAspHisPheThrIleThr 380
Db 1134 GGTGGCGTGCACCTTATCATGGCGGTGACGACGGGACCATTTTACCATCACC 1193
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1194 ACCCACTTGAGAGCAACCGGGCATCTCTGACCAACGAGAGGGTGTGGATTTTGAGGCC 1253
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAACACGACACACCTCTGATGAGTGAAGTGAACCAAGGAGGCCCTTTTGTGCTGAAG 1313
```

Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCACCATAGTGGTCCACGTGGAGGATGTGAATGAGGACCATGTGTGTT 1373
Qy 441 ValProProSerLysValValGluValGluGluIleProThrGlyGluProValCys 460
Db 1374 GTCCACCCCTCCAAAGTCGTGTAGGTCAGAGGGCATCCCACTGGGAGCCGTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACTGCAGAGAGCCCTGCACAGGAGATCAAAAGATCAGCTACCGCATCTTGAGA 1493
Qy 481 AspProIadGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCAGCGGGTGGCTAGCCATGGACCCAGAGAGTGGGAGGTGACAGTGGGGCACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGAGGATGAGCAGTTGTGAGGAAACACATCTATGAGTCATGTGTGGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp 540
Db 1614 ATGGACAATGGAGCCCTCCACACTGGCAGCGGAACCTTCTGTAACTGATGAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCNACGACCATGGCCAGTCCCTGAGCCCGTCAGATCACCATCTGCAACCAAGCCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGGCCACGTGTAACATCAGGACAGGAGCTGTCTCCACACCTCCCTTTCACAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 1794 GCCCAGCTCAGATGACTCAGACATCTCTGACGCGCAGAGGTCAACGAGGAGGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGTCTGTCCCTGGAAGATTCTCTGACGAGGATACATATGACGTGCACCTTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTACCATGCGCAACAAGAGCAGCTGACGCTGATCAGGGCCACTGTGTGCGACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTriPheIleLeuProValLeu 660
Db 1974 CATGGCCATGTCGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2034 GGGGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2093
Qy 681 ArgLysIleLysGluProLeuLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAAGAGAGCCCTCTACTCCAGAGATGACACCCGTGACACGCTTCTTAC 2153
Qy 701 TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCAAGAGGGGGTGGCGAAGAGACGACGACTATGACATCACCCAGCTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
Db 2214 GGTCTGAGGCGCGAGGGGTGTCTCCGAAATGACGTGGCACCACCATCATCCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspDluIleGlyAsnPheIleLeuGlu 760
Db 2274 ACACCCATGTACCGTCTAGCCAGCCACCCAGATGAAATCGCAACTTTATATTTAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGCGGCTAAACACAGACCCCGCCCGCCCTTACGACACCCCTCTTGGGTTC 2393

Qy 781 AspTyrGluGlySerGlySerAspAlaIleSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGCGACGGCTCCGACGCGGCTGAGCTCCCTGAGCTCCCTCCTCGCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCGCTTCAAGAACTGGCA 2513
Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 2514 GACATGTACGTGGTGGGAGGACGAC 2540
RESULT 3
US-09-880-107-3759
; Sequence 3759, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 3759
; LENGTH: 3171
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 X63629
US-09-880-107-3759
Alignment Scores:
Pred. No.: 0 Length: 3171
Score: 829.00 Matches: 829
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0
US-09-916-849A-1 (1-829) x US-09-880-107-3759 (1-3171)
Qy 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuGlnValCysTrpLeuGln 20
Db 54 ATGGGGCTCCCTCGTGGAGCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 113
Qy 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
Db 114 TCGCGGCTCCGAGCGCTCGCGGCGGTCTTCAGGAGGCTGAAGTACCTTGAGGCG 173
Qy 41 GlyGlyAlaGluGlnGluProGlyGluAlaLeuGlyLysValPheMetGlyCysProGly 60
Db 174 GGAGCGCGGAGCAGGAGCG 233
Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
Db 234 CAAGACCGAGCTCTGTGTAGCACTGATATGATGACTTCACTGTGCGGAATGGCGAGACA 293
Qy 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
Db 294 GTCCAGGAAAGAGGTCACTGAGGAAAGGAATCCATTGAAGATCTTCCATCCAAACGT 353
Qy 101 IleLeuArgArgHisLysLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
Db 354 ATCTTACGAAGACACAGAGAGATTGGGTGCTCTCAATATCTGCTCCCTGAAAAATGGC 413

QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
Db 414 AAGGTCCTTCCCTCCAGACAGTGAATCAGCTCAAGTCTAATAAAGATAGAGACACCAAG 473
QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal 160
Db 474 ATTTTCTACAGCATCAGCGGGCGGGGAGACAGCCCCCTCAGGGTGTCCTCGTGTA 533
QY 161 GluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
Db 534 GAGAAGAGACAGCGCTGTTGTTGTAATAAGCCACTGACCGGGAGGAGATGGCCAAG 593
QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 594 TATGAGCTCTTTGGCCACGCTGTGTGTCAGAGATGGTCCCTCAGTGAGGAGCCCATGAAC 653
QY 201 IleSerIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 654 ATCTCCATCATCGTGACCGCACAGATGACCAAGCCCAAGTTTACCCAGGACACCTTC 713
QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 714 CGAGGGAGTGCTTTAGAGGAGTGCCTACCGGTACTTCTGTGATGAGGTGACAGCCACA 773
QY 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
Db 774 GATGAGGATGATGCCATCTACACTACAAATGGGGTGGTTCCTTACTCCATCCATAGCCAA 833
QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 834 GAACCAAGAGACCCACAGACCTCATGTTCCAAATTCACCGGAGCACAGGCACCATCAGC 893
QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db 894 GTCATCTCCAGTGGCTGGACCGGGAANAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
QY 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluIleLeuAsp 320
Db 954 ACAGACATGATGGGACGGCTCCACCACCGCAGTGCAGTAGTGAGAGATCCTTGAT 1013
QY 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1014 GCCAATGACATGCTCCCATGTTTGACCCCAAGTACGAGGCCCATGTGCTCGAGAAAT 1073
QY 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1074 GCAGTGGGCCATGAGGTGCAGAGGCTGACGCTCACTGATCTGCACGCCCCCACTCACC 1133
QY 361 AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db 1134 GCGTGGCGTCCACTACCTTATCATGGCGGTGACGACGGGACCATTTTACCATCACC 1193
QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1194 ACCACCTCAGAGCAACCGAGGCATCTGACACACCGAGAGGTTTGGATTTTGAGGCC 1253
QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAAACAGACACACCTGTGACGTGAAGTACCAACAGAGCCCTTTTGTGCTGAAGCTC 1313
QY 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCACCTCCACAGCCCATAGTGTCCAGTGAGGATGTGAATGAGGACCTCTGTGTT 1373
QY 441 ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1374 GTCCACCTCCAAAGTCTGTGAGTCCAGAGGGCATCCCACTGGGGAGCCTGTGTGT 1433
QY 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTGCAGAGACCTGCAGAGGAGATCAAAAGATCAGCTACCGCATCCTGAGA 1493
QY 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500

RESULT 4

US-10-058-270A-125

Db 1494 GACCCAGCAGGGTGCTAGCTAGCCATGGACCCAGACAGTGGGAGGTGCACAGCTGTGGCACC 1553
QY 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGAGGATGAGCAGTGTGTGAGGAACAACATCATGAAGTCAATGGCTTGCC 1613
QY 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuThrLeuIleAsp 540
Db 1614 ATGGCAATGGAGAGCCCTCCACCACTGGCAGCGGGAACCTTCTGCTAACACTGATTGAT 1673
QY 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCACAGCACCATGGGCCCATGCTCCTGAGGCCCGTCCAGATCCATCTGCACCAAGCCCT 1733
QY 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGCGCCACGTCTGAAACATCACGACCAAGGACCTGTCTCCCCACACCTCCCTTTCCAG 1793
QY 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 1794 GCCCAGCTCACAGATGACTCAGACATCTACTGGACGGCAGAGGTCAACAGAGAGTGAC 1853
QY 601 ThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGGTCTTCTCCCTGAAGAAGTCTCTGAAGCAGGATACATATGACGTGCACCTTCT 1913
QY 621 LeuSerAspHisGlyAsnLysGluLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTGACCATGGCAACAAGAGCAGCTGACGGTGATCAGGGCCACTGTGTGCGACATGC 1973
QY 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu 660
Db 1974 CATGGCCATGTGCAAACTGCCCTGGACCTGGAAAGGAGGTTTCATCCTCCCTGTGCTG 2033
QY 661 GlyAlaValLeuAlaLeuPheLeuLeuValLeuLeuLeuValLargLysLys 680
Db 2034 GGGGTGTCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2093
QY 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2094 CGAAGATCAAGAGCCCTCTACTCCAGAGATGACCCCGTGACACAGCTCTTCTAC 2153
QY 701 TyrGlyGluGlyGlyGlyGluAspGlnAspTyrAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGCGAAGAGGGGGTGGGAGAGCAGGACTATGACATCACCAGCTCCACCGA 2213
QY 721 GlyLeuGluAlaAspProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
Db 2214 GGTCTGGAGGCGCAGCGCGAGTGTCTCCGCAATGACGTGGCACCACCATCATCCCG 2273
QY 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db 2274 ACACCCATGTACCGTCTTAGCCAGCCAAACCCAGATGAAATCGGCAACTTTATATTGAG 2333
QY 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAAGGGGCTTAAACAGACACCCACGCCCCCTCAGCACCCCTCTTGTGTGTC 2393
QY 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGGCAGCGGCTCCAGCGCGGCTCCCTGAGCTCCCTCACCTCTCCGCTCC 2453
QY 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCCGCTTCAAGAAGCTGCA 2513
QY 821 AspMetTyrGlyGlyGluAspAsp 829
Db 2514 GACATGTACGGTGGCGGGAGGACGAC 2540

654 ATCTCCATCATCGTGACCGACAGATGACCAAGCCCAAGTTTATCCAGGACACCTTC 713
Qy
221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db
714 CGAGGGAGTGTCTAGAGGAGTCCCTACCAAGGTACTTCTGTATGAGGATGACAGCCACA 773
Qy
241 AspGluAspAlaIleThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 260
Db
774 GATGAGATGATGCCATCTACCTACATCAATGGGGTGGTCTTACTCCATCCATGCCAA 833
Qy
261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db
834 GAACCAAGACCCACACGACCTCATGTTCCAAATTCACCGGACAGCACCATCAGC 893
Qy
281 ValIleSerSerGlyLeuAspArgGluValProGluThrThrThrThrThrThrThrThr 300
Db
894 GTCACTCTCCAGTGGCTGGACCGGGAAGTCCCTGAGTACACATGACCATCCAGGCC 953
Qy
301 ThrAspMetAspGlyAspGlySerThrThrThrThrThrThrThrThrThrThrThrThr 320
Db
954 ACAGCATGATGGAGCGGCTCCACCACCGCCAGTGGCAGTGGAGATCCTTGAT 1013
Qy
321 AlaAsnAspAlaProMetPheAspProGlnLysThrGluAlaHisValProGluAsn 340
Db
1014 GCCAATGACAAATGCTCCCAATGTTTGACCCCGAGAGTACGAGGCCCATGTGCTGAGAT 1073
Qy
341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db
1074 GCAGTGGGCCATGAGTGGAGGCTGACGGTCACTGATCTGACGCGCCCACTCACCA 1133
Qy
361 AlaTrpArgAlaThrThrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db
1134 GCGTGGCGTGCCACCTACCTATCATGGCGGTGACGACGGGACCACTTTTACCATCAC 1193
Qy
381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db
1194 ACCACCTTGAGACCAACCGGCTCTGACAAACAGGAGGGTTGGATTGAGGCC 1253
Qy
401 LysAsnGlnHisThrLeuThrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db
1254 AAAAACCCAGCACACCTGTAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1313
Qy
421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
Db
1314 CCAACCTCCACAGCACCATAGTGTCCAGTGGAGATGTAATGAGGACACCTGTGT 1373
Qy
441 ValProSerLysValValGluValGlnGlyIleProThrGlyGluProValCys 460
Db
1374 GTCCACCTCCAAAGTGGTGGTCCAGAGGACATCCCACTGGGAGCCCTGTGTGT 1433
Qy
461 ValThrThrAlaGluAspProAspLysGluAsnGlnLysIleSerThrArgIleLeuArg 480
Db
1434 GTCTACACTGCAGAAAGCCCTGACAGAGAGATCAAAAGATCACTACCCATCTTGAGA 1493
Qy
481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db
1494 GAACCCAGAGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1553
Qy
501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleThrGluValMetValLeuAla 520
Db
1554 CTCACCGTGGAGTGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1613
Qy
521 MetAspAsnGlySerProProThrThrThrThrThrThrThrThrThrThrThrThrThr 540
Db
1614 ATGACCAATGGAAGCCCTCCACCTCCACCTCCACCTCCACCTCCACCTCCACCTCCACCT 1673
Qy
541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db
1674 GTCAACGACCATGGCCCGGCTCCCTGAGCCCGTCCAGATCCCATCTGCAACCAAGGCCCT 1733
Qy
561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db
1734 GTGGGCCACGTGCTGAACATCAAGGACCTGTCTCCACACACCTCCCTCCCTTTCCAG 1793

581 AlaGlnLeuThrAspAspSerAspIleTrpThrAlaGluValAsnGluGlyAsp 600
Db
1794 GCCCAGCTCACAGATGACTCAGACATCTACTGGACGGCAGAGGTCAACGAGGAAGGTGAC 1853
Qy
601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrThrThrThrThrThrThr 620
Db
1854 ACAGTGGTCTTGTCTCCCTGAAGAAGTCTCCTGAAGCAGGATACATAGCGTGCACCTTCT 1913
Qy
621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db
1914 CTGCTGACCATGGCAACAGACAGCTGACGCTGATCAGGGCCACTGTGTGCGACTGC 1973
Qy
641 HisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValLeu 660
Db
1974 CATGCCATGTGCAAAACCTGCCCTGGAGCCCTGGAAAGAGGTTTCATCCTCCCTGTGCTG 2033
Qy
661 GlyAlaValLeuAlaLeuLeuPheLeuLeuLeuValLeuLeuLeuValArgLysLys 680
Db
2034 GGGGCTGTCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2093
Qy
681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db
2094 CGGAAGATCAAGGAGCCCTCTCTCTCCAGAGATGACACCCCTGACACCATCATCCCT 2153
Qy
701 TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db
2154 TATGCGGAAGAGGGGGTGGGAAGAGACGAGACTATGACATCACCGAGCTCCACCGA 2213
Qy
721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db
2214 GGCTCGAGGCGGAGCGGAGTGTCTCCGCAATGACGTGGCACCACCATCATCCCT 2273
Qy
741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db
2274 ACACCCATGTACCTCTCTAGCCAGCCAGCAACCATGAATCGGCAACTTTATTAATTGAG 2333
Qy
761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db
2334 AACCTGAGGCGGCTAACACAGACCCACACGCGCCCTACGACACCCCTCTTGGTGTTC 2393
Qy
781 AspTyrGluGlySerGlySerAspAlaAsnSerSerSerSerSerSerSerSerSerSer 800
Db
2394 GACTATGAGGCGGCGCTCCGACCGCGCTCCCTGAGCTCCCTCCTCCTCCTCCTCCTCCTC 2453
Qy
801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLysLeuAla 820
Db
2454 GACCAAGACCAAGATTACCATTTCTGACGATGGGCGAGCCGCTTCAGAAAGCTGGCA 2513
Qy
821 AspMetTyrGlyGlyGluAspAsp 829
Db
2514 GACATGTACGCTGGCGGAGGAGGAC 2540

RESULT 6
US-10-172-118-669
; Sequence 669, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380, 770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699


```
Db 1194 ACCACCTGAGAGCAACCGGCGATCCCTGACAAACAGGAGGTTTGATTTTGGAGCC 1253
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAACAGCAGCACACCTGTACGTTGAAGTGACCAACAGAGGCCCTTTTGTGCTGAAGCTC 1313
Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCCACCATAGTGGTCCACGTCGAGGATGTGAATGAGGACCTGTGTTT 1373
Qy 441 ValProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
Db 1374 GTCCCAACCTCCAAAGTCGTTGAGGTCAGGAGGCGATCCCCACCTGGGAGCGCTGTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTGCAGAGACCCCTGACAGGAGAAATCAAAGATCATGCTACCGCATCTGTGAGA 1493
Qy 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCAGCAGGTCGCTAGCCATGGACCCAGACAGTGGCGAGGTCCACACTGTGGGCACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTCAGGATGAGCAGTTTGTGAGGAAACAACATCATCAAGTCATGCTCTGGCC 1613
Qy 521 MetAspAsnGlySerProThrThrGlyThrGlyThrLeuLeuLeuThrIleLeuAsp 540
Db 1614 ATGGCAATGGAGGCCCTCCCACTGGCAGCGGAACCTCTCTGCCACACCTCCCTCTTCAG 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGTGCCCGCAGTCCTGAGCCCGCTCAGATCACCATCTGCAACCAAGCCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGCGCCAGTCGTGAACATCAGGACAGGACCTGTCTCCCAACACCTCCCTCTTCAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGlyAsp 600
Db 1794 GCCAGCTCACAGATGACTCAGACATCTACTGGACGCGAGAGGTCAACGAGAAAGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGTCTTGTCCCTGGAAGATCTCTGAAGCAGGATACATATCAGTCGACCTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CHTCTGACCATGGCAACAAAGACGACGTGACGGTGATCAGGCGCACCTGTGCGCACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTyrLysGlyPheIleLeuProValLeu 660
Db 1974 CATGGCCATGTGCAAACTCGCCCTGGACCTGGAAAGGAGGTTTCATCTCCCTGTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuPheLeuLeuValLeuValLeuLeuValArgLysLys 680
Db 2034 GGGCGTCTCTGTCTGTCTGTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2093
Qy 681 ArgLysIleLysGluProLeuLeuProGluAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAAGAGCCCTCTACTCCAGAGATGACACCCGTCGACAACTCTTCTAC 2153
Qy 701 TyrGlyGluGluGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCGAAGAGGGGTGGCGAAGAGGACAGGACTATGACATCACCAGCTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro 740
Db 2214 GGTCTGGAGGCCAGGCCGAGGTGTCTTCCCAATGACGTGGCACCAACCAATATCCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleLeuGlu
```

```
Db 2274 ACACCATGTACGTCCTTAGCCAGCCAAACCCAGATGAAATCGGCACACTTTATTAATTGAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGCGGCTTAACACAGACCCACAGCCCGCCCTACGACACCTCTTGTGTGTTT 2393
Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGCGAGCGGCTCCGACCGCGCTCCCTGAGCTCCCTCAGCTTCCCTCCGCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGACGAGTGGGCGACCGCTTCAAGAAGCTGGCA 2513
Qy 821 AspMetTyrGlyGlyGluAspAsp 829
Db 2514 CACATGTACGTCGGGCGGAGGACGAC 2540

RESULT 8
US-10-158-123-2
; Sequence 2, Application US/10158123
; Publication No. US2003019406A1
; GENERAL INFORMATION:
; APPLICANT: Reinhard, Christoph
; APPLICANT: Klinger, Julie
; APPLICANT: Jefferson, Ann
; APPLICANT: Escobedo, Jaime
; APPLICANT: Randazzo, Filippo
; APPLICANT: Winter, Jill
; APPLICANT: Goodson, Robert
; APPLICANT: Qi, Weimin
; TITLE OF INVENTION: P-Cadherin as a Target for Anti-Cancer
; FILE OF INVENTION: Therapy
; FILE REFERENCE: 35784/255994
; CURRENT APPLICATION NUMBER: US/10/158,123
; PRIOR FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: 60/294,225
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 3171
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-158-123-2

Alignment Scores:
Pred. No.: 0 Length: 3171
Score: 829.00 Matches: 829
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 15 Gaps: 0

US-09-916-849A-1 (1-829) x US-10-158-123-2 (1-3171)
Qy 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln 20
Db 54 ATGGGGCTCCCTCGTCGTGGACCTCTCGCTCTCTCTCTCTCCAGGTTTGTGCTGCTGAG 113
Qy 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
Db 114 TGGCGGCTCCCGAGCGCTGCGGCGGTCTTCAGGGAGGCTGAAGTGACCTTGGAGGCG 173
Qy 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
Db 174 GGAGGCGGAGGAGGAGCGCGCCAGCGCTGGGGAAGATTTCATGGCTGCCCTGGG 233
Qy 61 GlnGluProAlaLeuPheSerThrAspAsnAspAspPheThrValArgAsnGlyGluThr 80
Db 234 CAAGAGCCAGCTCTGTATTAGCACTGATAATGATGACTCATCTGTGCGGAATGCGGAGCA 293
Qy 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
```

Db 294 GTCCAGGAAGAGGTCACCTGAGGAAAGAAATCCATTCAGATCTTCCCATCAAAAGCT 353
Qy 101 ILeuAArgHISlyArgAspTrpValAlaProIleSerValProGluAsnGly 120
Db 354 ATCTTACGAGACACAGAGAGATGGGTGGTTCCTCAATATCTGTCTGCTGAAATGGC 413
Qy 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
Db 414 AAGGGTCCCTTCCCCAGAGACTGATCAGCTCAAGTCTAATAAGATAGACACCAAG 473
Qy 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal 160
Db 474 ATTTCTACAGCATCACGGGCGGGGCGAGACAGCCCTGAGGGTGTCTTCGCTGTA 533
Qy 161 GluLysGluThrGlyTrpLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
Db 534 GAGAGGAGACAGCTGGTGTGGTGGTAAATAGCCCACTGGACCGGAGGAGATTGCCAAG 593
Qy 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 594 TATGAGCTCTTTGGCCACGCTGTGTGAGAGATGGTGCCTCAGTGGAGGACCCCATGAAC 653
Qy 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 654 ATCTCCATCATCTGACCGACCGAGATGACCAAGCCCAAGTTTACCAGGACACCTTC 713
Qy 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 714 CGAGGGAGTGTCTAGAGGGAGTCTTACCAGGTACTTCTGTGATGCAGGTGACAGCCACA 773
Qy 241 AspGluAspAspAlaIleTyrThrAsnGlyValAlaValAlaValTyrSerIleHisSerGln 260
Db 774 GATGAGGATGATGCCATCTACCTACCTCAATGGGGTGGTGTCTTACTCCATCCATAGCCAA 833
Qy 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 834 GAACCAAGAGACCCACACGACCTTCACTGTTACANTTCAACCGAGACAGGACCATCAGC 893
Qy 281 ValIleSerSerGlyLeuAspAArgLysValProGluTyrThrLeuThrIleGlnAla 300
Db 894 GTCATCTCCAGTGGCTGGACCGGAAAAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
Qy 301 ThrAspMetAspGlyAspGlySerThrThrThrAlaValAlaValValGluIleLeuAsp 320
Db 954 ACAGACATGATGGGACGGCTCCACACCGCCGAGTGGCAGTAGTGGAGATCTTGTAT 1013
Qy 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1014 GCCAATGACAATGTCTCCATGTTTGACCCCCAGAGATACGAGGCCCATGTGCTGAGAAAT 1073
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1074 GCAGTGGGGCCATGAGGTGCAGAGCTGACCGTCACTGATCTGACGCCCCCAACTCACCA 1133
Qy 361 AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
Db 1134 GCGTGGGTGTCACCTACTATCATGGGCGGTGACGACGGGACCATTTTACCATCAC 1193
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1194 ACCCACCTGAGAGCAACCGGGCATCTTGACAAACAGGAGGGTTTGGATTTTGAGGCC 1253
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAACCCAGACACCCCTGACCTTGAAGTGAACCAACGAGGCCCTTTTGTGCTGAAGCTC 1313
Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCCACCATAGTGTCCAGTGGAGGATGTGAATGAGGCACCTGTGT 1373
Qy 441 ValProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460

Db 1374 GTCCCACTCCAAAGTCGTGTGAGGTCAGAGGGCATCCCCACTGGGAGGCTGTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTGACAGAGACCTGACAAAGGAGAAATCAAAAGATCAGCTACCGCATCTGAGA 1493
Qy 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCACGACGAGGTGGCTAGCCATGGACCCAGACAGTGGGAGGTCCACAGCTGTGGGCACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCACCGCTGAGGATGAGCAGTTTGTGAGGAACAACATCTATGAAGTCATGGCTTGGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeuAsp 540
Db 1614 ATGGACAATGGAAGCCCTCCACACCTGGCAGCGGAACCTTCTGTAAACACTGATTGAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGGCCCGCCAGTCCCTGAGCCCGTCAGATCACCATCTGCAACCAAGCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGGCCACGTGTCTGAACATACGACAGAGACCTGTCTCCACACACCTCCCTTCCAG 1793
Qy 581 AlaglnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGluValAsp 600
Db 1794 GCCAGCTCAGATGATCAGATCTCTCTGACCGCAGAGTCAACGAGGAAGGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGTCTTGTCTCCCTGAGAAGTCTCTGAGCAGGATACATATGACGTGCACCTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTGACCATGCAAAAGAGCAGCTACGGTGTATCAGGCGCACTGTGTGGACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTrpLysGlyGlyPheIleLeuProValLeu 660
Db 1974 CATGCCATGTCCAAACTGCCCTGGACCTCGAAGAGAGGTTCATCTCCCTGTGTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuLeuPheLeuLeuLeuValLeuLeuLeuValArgLysLys 680
Db 2034 GGGGCTGTCTGGTCTGTCTCTCTCTGTGTGTGTCTTTTGTGTGTGAGAAAGAG 2093
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAAGGAGCCCTCTCTACTCCAGAGATGACACCCGTGACACGCTTCTTAC 2153
Qy 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGGAGAGGGGGGTGGCGAAGAGACCCAGCACTATGACATCACCCAGCTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValIleuArgAsnAspValAlaProThrIleIlePro 740
Db 2214 GGTCTGGAGGCCGCGGAGGTCTCTCGCAATGACGTGGGACCAACCATCATCTCCG 2273
Qy 741 ThrProMetTyrAspProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db 2274 ACACCCATGTACCTCTCTAGCCGACCAACCCAGATGAATCGGCAACTTATATATGAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGGGGGTAAACACAGACCCACGCGCCCTACGACACCTCTTGGTGTTC 2393
Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGCGGGCTCCGACGCGGCTCTCTGAGCTCCCTCACCTCTCCGCGCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLeuAla 820
Db 2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGCGAGCCGCTTCAAGAGACTGGCA 2513

QY 821 AspMetTyrGlyGlyGluAspAsp 829
Db 2514 GACATGACGTGGCGGGAGGACGAC 2540
RESULT 9
US-10-295-027-782
; Sequence 782, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 782
; LENGTH: 3171
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-782
Alignment Scores:
Pred. No.: 0 Length: 3171
Score: 829.00 Matches: 829
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 16 Gaps: 0
US-09-916-849a-1 (1-829) x US-10-295-027-782 (1-3171)
QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln 20
Db 54 ATGGGGCTCCCTCGTGAGACCTCTCGGGTCTCTCCTCTTCAGGTTTGGCTGGCAG 113
QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
Db 114 TGGCGGGCTCCCGAGCGGCGGGGGTCTTCAGGGAGGCTGAAGTGACCTTGGAGGCG 173
QY 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60

Db 174 GGAGCGCGGAGCAGGAGCCCGCGCGTGGGAAAAGTATTTCATGGGCTGCCCTGGG 233
QY 61 GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr 80
Db 234 CAAGAGCCAGCTCTGTTTAGCACTGATATGATGATCTCCTGCTGGGAATGCCGAGACA 293
QY 81 ValGlnGluArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
Db 294 GTCCAGGAAAGAGGTCACTGAAGGAAAGGAATCCATTGAAGATCTCCCATCCAAAGCT 353
QY 101 IleLeuArgArgHisLysArgAspTrpValValAlaProLleSerValProLleGlnGly 120
Db 354 ATCTTACGAGACACAAGAGATTGGGTGGTGTCTCAATATCTGTCCCTGAAATGGC 413
QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
Db 414 AAGGGTCCCTTCCCGCAGAGACTGAATCAGCTCAAGTCTAATAAAGATAGACACCAAG 473
QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
Db 474 ATTTTCTACAGCATCACGGGGCGGGGCGAGACACCCCTGAGGGTGTCTTCCTCTGTA 533
QY 161 GluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
Db 534 GAGAAGGAGACAGGCTGGTGTGTTGATTAAGCCACTGGACCGGAGAGATTGCCAAG 593
QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
Db 594 TATGAGCTCTTTGGCCACGCTGTGTGAGAGATGGTGCCTCAGTGGAGAGCCCCATGAAC 653
QY 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
Db 654 ATCTCCATCATCGTGACCGACGAGATGACCAAGCCCAAGTTTACCAGGACACCTTC 713
QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
Db 714 CGAGGGAGTGTCTTAGAGGAGTCTTACAGGTACTCTGTGATGACAGTGCAGCCACA 773
QY 241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
Db 774 GATGAGATGATGCCATCTACCTTACANTGGGTGGTGTCTTACTCCTCATAGCCAA 833
QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
Db 834 GAACCAAGGACCCACACGACTCATGTTTCAATTCACCGGAGCAGACGACCATCAGC 893
QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
Db 894 GTCATCTCCAGTGGCTGGACCGGAAAAAGTCCCTGAGTACACACTGACCATCCAGGCC 953
QY 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValGluIleLeuAsp 320
Db 954 ACAGACATGGATGGGACGGCTCCACCAACGAGTGGCAGTAGTGGAGATCCTTGAT 1013
QY 321 AlaAsnAspAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
Db 1014 GCCAATGCAATGCTCCCATGTTTACCCCCCAGAGTAGCAGGCCCCATGCTGCTGAGAAT 1073
QY 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1074 GCAGTGGGCCATGAGGTGACAGGCTGACGGTCTACTGATCTGGACGCCCCCACTCACA 1133
QY 361 AlaTrpArgAlaThrTyrLeuIleMetGlyLysAspGlyAspGlyHisPheThrIleThr 380
Db 1134 CGGTGGCGTGCCACTTATCATGGCGGTGACGACGGGACCATTTTACCATCACC 1193
QY 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
Db 1194 ACCCACCTGAGACACACGAGGATCTTGACACACGAGGAGGGTTTGGATTGAGGCC 1253
QY 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420

Db 1254 AAAAACACGACACCCCTGTACGTTGAAGTGACCAACGAGGCCCTTTTGTGCTGAAGCTC 1313
Qy 421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCCACCATAGTGGTCCACGTGGAGATGTGAATGAGGACCTGTGTTT 1373
Qy 441 ValProProSerLysValValGluValGlnGluGlyLeuProThrGlyGluProValCys 460
Db 1374 GTCCACCCCTCCAAAGTCGTTGAGTCCAGAGGCGATCCCACTGGGAGCGCTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTGCAGAGAGCCCTGCAAGAGAAATCAAAAGATCAGCTACCGCATCTTGAGA 1493
Qy 481 AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCAGCAGGGTGGCTAGCCATGGACCCAGACAGTGGCGAGCTGAGCTGGGCACC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGGAGTGCAGCATTTGTGAGGAACAACATCTATGAAGTCATGGTCTTGGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuLeuLeuLeu 540
Db 1614 ATGACAAATGGAACCTCTCCACCACTGGCAGCGGAACCTTCTGCTAACACTGATGAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGGCCAGTCCCTGAGCCCGTCAGATCACTCTGCAACCAAGGCCCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAsnLeuSerProHisThrSerProPheGln 580
Db 1734 GTGGCCACCGTGGTGAACATCACGGAAGACCTGTCTCCACACCTCCCTTTCCAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 1794 GCCAGCTCACAGATGACTCAGACATCTACTGGAGCGCAGAGGTCAACGAGGAAGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGTCTTGTCCCTGAAGAGTCTCTGAAGCAGGATACATATGACGTGCACCTTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTGACCATGGCAACAAAGACAGCTGACGTGTATCAGGCCCATCTGTGGACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTrpLysGlyPheIleLeuProValLeu 660
Db 1974 CATGCCATGTGAAACCTGCCCTGGACCTCTGAAAGAGGAGTTTTCATCTCCCTGTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuPheLeuLeuLeuValLeuLeuLeuValAlaGlyLysLys 680
Db 2034 GGGGCTGTCTGGCT 2093
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAAGAGGCCCTCTACTCCCAAGATGACACCCCGTGAACACGCTCTCTAC 2153
Qy 701 TyrGlyGluGluGlyGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCGAAGAGGGGGGGGGGAGAGAGACGAGACTATGACATCACCCAGCTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db 2214 GGTCTGGAGGCCAGGCCGAGGTGTCTCTCGCAATGACGTGGCAACCAACCATCATCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db 2274 ACACCCATGTACCGTCTCTAGCCAGCCCAACCCAGATGAAATCGGCACACTTTATTAAT 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGGGGCTTAAACAGACACCCCGCCCTACGACACCTCTCTGTGTCTC 2393

Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGCGAGCGGCTCCGACGCGGCTCCCTGAGCTCCCTCAGCTCCTCCGCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLeuAla 820
Db 2454 GACCAAGACCAAGATTCAGATTATCTGAACGATGGGCGAGCCGCTTCACAGAGCTGCA 2513
Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 2514 GACATGTACGCTGGCGGAGGAGCAGC 2540
RESULT 10
US-10-295-027-799
; Sequence 799, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT FILING DATE: 2002-11-13
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 799
; LENGTH: 3171
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-799
Alignment Scores:
Pred. No.: 0 Length: 3171
Score: 829.00 Matches: 829
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 16 Gaps: 0
US-09-916-849a-1 (1-829) x US-10-295-027-799 (1-3171)
Qy 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuGlnValCysTrpLeuGln 20

Db	54	ARGGGGCTCCCTCGTGGACCTCTCGCGTCTCCCTCTCCAGGTTGCTGGCTGCAG	113
Qy	21	CysAlaIaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla	40
Db	114	TGCGCGGCTCCGAGCGCTCGCGGCGGTCTTCAGGAGGCTGAAGTGCACCTTGGAGGCG	173
Qy	41	GlyGlyAlaGluGlnProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly	60
Db	174	GGAGCGGGAGCAGAGACCGCGGCGAGCGCTGGGGAAAGTATTATCATGGGCTGCCCTGGG	233
Qy	61	GlnGluProAlaLeuPheSerThrAspAsnAspPheThrValArgAsnGlyGluThr	80
Db	234	CAAGAGCCAGCTCTGTTAGCACTGATTAATGATGACTTCATCTGTCGGAAATGGCGAGACA	293
Qy	81	ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg	100
Db	294	GTCCAGGAAAGAGTCTCAAGAGAAAGGAATCCATTGAAGATCTTCCCATCCAAACGT	353
Qy	101	IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly	120
Db	354	ATCTTACAGACACACAGAGAGATTGGTGTGTCTCAATATCTGTCCCTGAAATGGC	413
Qy	121	LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys	140
Db	414	AAGGTCCTCTCCCCAGAGACTGAATCAGCTCAAGTCTAATATAAGATAGAGACACCAAG	473
Qy	141	IlePheTyrSerIleThrGlyProGlyAlaAspSerProProGluGlyValPheAlaVal	160
Db	474	ATTTTCTACAGCATCACGGGCGGGGGCAGACGCCCTCGTAGGGGTCTCTCGCTGTA	533
Qy	161	GluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys	180
Db	534	GAGAAGGAGACAGGCTGTTGTTGTAATAAGCCACTGGACCGGGAGAGATTGCCAAG	593
Qy	181	TyrGlnLeuPheGlyHisAlaValSerGlnAsnGlyAlaSerValGluAspProMetAsn	200
Db	594	TATCAGCTCTTTGGCCACGCTGTGTACAGAAATGGTGCTCAGTGGAGACCCCATGAAC	653
Qy	201	IleSerIleIleValThrAspGluAsnAspHisLysProLysPheThrGlnAspThrPhe	220
Db	654	ATCTCCATCATCTGTGACCGCACAGATGACACAGCCCAAGTTTACCAGACACCTTC	713
Qy	221	ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr	240
Db	714	CGAGGAGTGTCTTAGAGGAGTCTCTACAGGTACTTCTGTGATCGAGTGACAGCCACA	773
Qy	241	AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln	260
Db	774	GATGAGGATGATGCCATCTCACCTACAAATGGGGTGTGTTACTTCCATCCATAGCGAA	833
Qy	261	GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer	280
Db	834	GAACCAAGGACCCACACGACTCATGTTCAATTCAACCGGAGACAGGCACCATCAGC	893
Qy	281	ValIleSerSerGlyLeuAspArgGlnLysValProGluTyrThrLeuThrIleGlnAla	300
Db	894	GTCATCTCCAGTGGCTGGACCGGAAAGTCCCTGAGTACACACATGACCATCCAGGCC	953
Qy	301	ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluIleLeuAsp	320
Db	954	ACAGCATGGATGGGACGGCTCCACCACCGCAGTGGCAGTAGTGGAGATCCTTGAT	1013
Qy	321	AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn	340
Db	1014	GCCAAATGACAAATGCTCCCATGTTTGACCCCGAGAAAGTACGAGGGCCCATGTGCCTGAGAAT	1073
Qy	341	AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro	360
Db	1074	GCATGGGCCCATGAGGTGCAGAGGCTGACGCTCACTGATCTGGAGCGCCCACTACACA	1133
Qy	361	AlaTrpArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr	380

Db	1134	CGTGGCGTCCACCTACCTTATCATGGGGGTGCACGCGGGACCAATTTTACCATCACC	1119
Qy	381	ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla	400
Db	1194	ACCACCTCGAGAGCAACCGAGGCATCCTGACACACGAGGAGGTGTGGATTGTAGGCC	1253
Qy	401	LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu	420
Db	1254	AAAAACACGACACCCCTGTAGTTGAAGTGAACCAACGAGGCCCTTTGTGCTGAAGCTC	1313
Qy	421	ProThrSerThrAlaThrIleValIleValHisValGluAspValAsnGluAlaProValPhe	440
Db	1314	CCACCTCCACAGCCACCATAGTGTCTCAAGTGGAGGATGTGAATGAGGCACCTGTGTT	1373
Qy	441	ValProProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys	460
Db	1374	GTCCACCCCTCAAAGTCGTTGAGTCCAGAGGCGCATCCCACTGGGAGCGCTGTGTGT	1433
Qy	461	ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg	480
Db	1434	GTCTACACTGCAGAAGACCCCTGACAGAGAGATCAAAAGATCACTACCCGATCCTTGAGA	1493
Qy	481	AspProAlaGlyTyrLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr	500
Db	1494	GACCCAGCAGGTCGCTAGCCATGGACCCACAGACAGTGGCAGGTCAACAGTGTGGGCACC	1553
Qy	501	LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla	520
Db	1554	CTCAGCCGTGAGGATGAGCAGTTGTGAGGAAACAACATCTATGAAGTCATGGCTTGGGCC	1613
Qy	521	MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuIleAsp	540
Db	1614	ATGGACAATGGAAGCCCTCCACACACTGGCACGGGAACCCCTTCTGCTAACACTGATTGAT	1673
Qy	541	ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro	560
Db	1674	GTCAACGACCATGGCCCGAGTCCTCGACGCCCGTCAGATCCACATCTGCACACCAAGGCCCT	1733
Qy	561	ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln	580
Db	1734	GTGGCCACCTGCTGAACATCACGGACAAGGACCTGTCTCTCCACACCTCCCTTCCAG	1793
Qy	581	AlaGlnLeuthrAspAspSerAspIleTyrTyrThrAlaGluValAsnGluGlyAsp	600
Db	1794	GCCCAGCTCACAGATGACTCAGACATCTACTGACGCGCAGAGGTCAACGAGGAGGTGAC	1853
Qy	601	ThrValValLeuSerLeuLysPheLeuLysGlnAspThrTyrAspValHisLeuSer	620
Db	1854	ACAGTGTCTTGTCCCTCGAAGAAGTTCTCTGAAGCAGGATACATATGACGTGCACCTTCT	1913
Qy	621	LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys	640
Db	1914	CTGTCTCACCATGGCAACAAGAGCAGCTCACGGTGATCAGGGCCACTGTGTGCGACTGC	1973
Qy	641	HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu	660
Db	1974	CATGGCCATGTGCAAAACCTCCCTGGACCCCTGAAAGGAGTTTCATCTCCCTGTGCTG	2033
Qy	661	GlyAlaValLeuAlaLeuLeuPheLeuLeuValLeuLeuLeuValArgLysLys	680
Db	2034	GGGCTGTCTGCTGTGCTGTTCCCTCTGCTGTGCTGCTTTTGTGTGAGAGAAG	2093
Qy	681	ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr	700
Db	2094	CGGAAGATCAAGGAGCCCTCTACTCCCAAGAAGATGACCCCGTGACACGCTTCTTCTAC	2153
Qy	701	TyrGlyGluGluGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg	720
Db	2154	TATGGCGAAGAGGGGGTGGCGAAGAGGACGAGCATATGACATCAACCCAGCTCCACCGA	2213
Qy	721	GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleLeuPro	740
Db	2214	GGTCTGAGGCGCAGCCGAGGTGGTTCTCCGCAATGACGTGGCAACAACCATCATCCCCG	2273

QY 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
 Db 2274 ACACCCATGTACCGCTCCTAGCCAGCCACCCAGATGAATCGCAACITTTATATTGAG 2333
 QY 761 AsnLeuIysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
 Db 2334 AACCTGAAGCGGCTAAACACAGACCCACAGCCCGCCCTACGACACCCCTCTTTGGTGTTC 2393
 QY 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
 Db 2394 GACTATGAGGCGAGCGCTCCGAGCGCGCTCCCTGAGCTTCTTACCTTCTCCGCTCC 2453
 QY 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheIysLeuAla 820
 Db 2454 GACCAAGACCAAGATTACGATTATCTACAGAGTGGGCGAGCGCTTCAAGAGCTGGCA 2513
 QY 821 AspMetTyrGlyGlyGlyGluAspAsp 829
 Db 2514 GACATGTACGCTGGCGGGAGGACGAC 2540
 RESULT 11
 US-10-295-027-862
 ; Sequence 862, Application US/10295027
 ; Publication No. US20030232350A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Afar, Daniel
 ; APPLICANT: Aziz, Natasha
 ; APPLICANT: Ginsberg, Wendy M.
 ; APPLICANT: Gish, Kurt C.
 ; APPLICANT: Glyne, Richard
 ; APPLICANT: Hevezi, Peter A.
 ; APPLICANT: Mack, David H.
 ; APPLICANT: Murray, Richard
 ; APPLICANT: Watson, Susan R.
 ; APPLICANT: Eos Biotechnology, Inc.
 ; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
 ; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
 ; FILE REFERENCE: 018501-012500US
 ; CURRENT APPLICATION NUMBER: US/10/295,027
 ; PENDING FILING DATE: 2002-11-13
 ; PRIOR APPLICATION NUMBER: US 09/663,733
 ; PRIOR FILING DATE: 2000-09-15
 ; PRIOR APPLICATION NUMBER: US 60/350,666
 ; PRIOR FILING DATE: 2001-11-13
 ; PRIOR APPLICATION NUMBER: US 60/335,394
 ; PRIOR FILING DATE: 2001-11-15
 ; PRIOR APPLICATION NUMBER: US 60/332,464
 ; PRIOR FILING DATE: 2001-11-21
 ; PRIOR APPLICATION NUMBER: US 60/334,393
 ; PRIOR FILING DATE: 2001-11-29
 ; PRIOR APPLICATION NUMBER: US 60/340,376
 ; PRIOR FILING DATE: 2001-12-14
 ; PRIOR APPLICATION NUMBER: US 60/347,211
 ; PRIOR FILING DATE: 2002-01-08
 ; PRIOR APPLICATION NUMBER: US 60/347,349
 ; PRIOR FILING DATE: 2002-01-10
 ; PRIOR APPLICATION NUMBER: US 60/355,250
 ; PRIOR FILING DATE: 2002-02-08
 ; PRIOR APPLICATION NUMBER: US 60/356,714
 ; PRIOR FILING DATE: 2002-02-13
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 1386
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 862
 ; LENGTH: 3171
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-295-027-862
 Alignment Scores:
 Pred No.: 0 Length: 3171
 Score: 829 Matches:

Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 16 Gaps: 0
 US-09-916-849A-1 (1-829) x US-10-295-027-862 (1-3171)
 QY 1 MetGlyLeuProArgGlyProLeuAlaSerLeuLeuLeuGlnValCysTrpLeuGln 20
 Db 54 ATGGGGCTCCCTCGTGGACCTCTCGCGTCTCTCTCTCTCTCCAGGTTTCTGCTGCTCAG 113
 QY 21 CysAlaAlaSerGluProCysArgAlaValPheArgGluAlaGluValThrLeuGluAla 40
 Db 114 TCGCGGCTCCGAGCGCTGCGGCGGTCTTCAGGAGGCTGAGTGACCTTGGAGGCG 173
 QY 41 GlyGlyAlaGluGlnGluProGlyGlnAlaLeuGlyLysValPheMetGlyCysProGly 60
 Db 174 GAGGCGCGGAGCAGGAGCGCGCGCTGGGAAAGTATTTCATGGGTGGCTGGCTGGG 233
 QY 61 GlnGluProAlaLeuPheSerThrAspAspAspPheThrValArgAsnGlyGluThr 80
 Db 234 CAAGAGCCAGCTCTGTTAGCACTGATTAATGATGACTTCACTGTGCGGATGGCGACACA 293
 QY 81 ValGlnGluArgArgSerLeuLysGluArgAsnProLeuLysIlePheProSerLysArg 100
 Db 294 GTCCAGGAAAGAGGTCTACTGAGGAAAGGAATCCATTGAAGATCTTCCCATCCAAAGCT 353
 QY 101 IleLeuArgArgHisLysArgAspTrpValValAlaProIleSerValProGluAsnGly 120
 Db 354 ATCTTTACGAAGACACACAGAGAGATTGGGTGTTCTCTCAATATCTGCTCCCTGAAATGGC 413
 QY 121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
 Db 414 AAGGTCTCTTCCCGCAGAGACTGNACTCAAGTCTAATTAAGTAGAGACACCAAG 473
 QY 141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGlyValPheAlaVal 160
 Db 474 ATTTTCTACAGCATCACGGGGCCGGGGCAGACAGCCCCCTGAGGGTGTCTTCGCTGTA 533
 QY 161 GluLysGluThrGlyTrpLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
 Db 534 GAGAAGGAGACAGGCTGTTGTTGTTGNAATGAAGCCACTGGACCGGAGGAGATTGCCAAG 593
 QY 181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
 Db 594 TATGAGCTCTTTGGCCACGCTGTGTGAGAGAATGGTGTCTCAGTGGAGGAGCCCATGAAC 653
 QY 201 IleSerIleIleValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
 Db 654 ATCTCCATCATCGTGACCGCAGCAATGACCACAGCCCAAGTTTACCAGGACACCTTC 713
 QY 221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
 Db 714 CGAGGAGTGCTTTAGAGGGAGTCTTACAGGTACTTCTGTGATGCAAGTGCAGGCCACA 773
 QY 241 AspGluAspAspAlaIleTyrThrAsnGlyValValAlaTyrSerIleHisSerGln 260
 Db 774 GATGAGGATGATGCCATCATACCTACCAATGGGTGTGTCTTACTCATCATAGCCAA 833
 QY 261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
 Db 834 GAAACCAAGGACCCACACAGCTCATGTTCATTCACCGGAGCAGCAGCCATCAGC 893
 QY 281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGluAla 300
 Db 894 GTCACTCCAGTGGCTCGACCGGAAAAAGTCCCTGAGTAGTACACACTGACCATCCAGGCC 953
 QY 301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluIleLeuAsp 320
 Db 954 ACAGACATGGATGGGACGGCTCCACCCACCGCAGTGGCAGTAGTAGGAGATCTTGTAT 1013
 QY 321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340

Db 1014 GCCAATGACAATGCTCCCATGTTTGAACCCAGAAAGTACGAGGCCCATGTGCTCGAAT 1073
Qy 341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
Db 1074 GCATGGCCCATAGGTTGAGAGGCTACGGCTACTGATCTGGAGGCCCACTACCA 1133
Qy 361 AlaTrpArgAlaThrTyrLeuIleMetGlyValAspAspGlyAspHisPheThrIleThr 380
Db 1134 GCGTGGCGTGCACCTACCTTATCATGGCGGTGACGACGGGACCACTTTTACCATCAC 1193
Qy 381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgIysGlyLeuAspPheGluAla 400
Db 1194 ACCACCCCTGAGAGCAACACCGGGCATCTGACACACCGAAGGGTTGGATTGAGGCC 1253
Qy 401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
Db 1254 AAAAAACGACACACCTGTAGTTGAAGTACCAACGAGGCCCTTTTGTGCTGAAGCTC 1313
Qy 421 ProThrSerThrAlaThrIleValValHisValGluAspValAsnGluAlaProValPhe 440
Db 1314 CCAACCTCCACAGCCACCATAGTGTCCAGCTGAGGATGTGAATGAGGCACCTGTGTTT 1373
Qy 441 ValProProSerLysValValGluValGlnGlyIleProThrGlyGluProValCys 460
Db 1374 GTCCACCCCTCCAAAGTCGTTGAGTCCAGAGGGCATCCCCACTGGGAGCCCTGTGTGT 1433
Qy 461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
Db 1434 GTCTACACTCAGAAAGACCTGACAGAGAGATCAAAAGATCAGCTACCGCATCTTGAGA 1493
Qy 481 AspProAlaGlyTrpLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
Db 1494 GACCCAGCGGTGGCTAGCATCGACCCAGAGGGGAGTGGGAGTGCACAGTGGGGCAC 1553
Qy 501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
Db 1554 CTCGACCGTGAGGATGAGCAGTTTGTGAGGACAAACATCATGATGATGCTTGGCC 1613
Qy 521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuThrLeuIleAsp 540
Db 1614 ATGGACATCGAAGCCCTCCACCATCGGCACCGGAAACCTTCTGTAAACACTGATTGAT 1673
Qy 541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
Db 1674 GTCAACGACCATGGCCAGCTCCCTGAGCCCGTCAGATCAACATCGACCAACCAACCT 1733
Qy 561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
Db 1734 GTGCGCCACGTGTGAACATCAGGACACAGGACCTGTCTCCACACCTCCCTTCCAG 1793
Qy 581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db 1794 GCCAGCTCAGATGACTCAGACATCTCTGACGCGCAGAGTCAACGAGGAAGTGAC 1853
Qy 601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db 1854 ACAGTGTCTTGTCCCTGGAAGAGTTCCTGAAGCAGGATACATATGACGTGCACCTTTCT 1913
Qy 621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db 1914 CTGTCTACCATGGCAACAAAGACGAGTCAACGCTGATCAGGGCCACTGTGTGCGACTGC 1973
Qy 641 HisGlyHisValGluThrCysProGlyProTyrTrpLysGlyGlyPheIleLeuProValLeu 660
Db 1974 CATGGCCATCTGAAACCTGCCCTGACCCCTGAAAGAGGTTTCATCTCCCTGTGCTG 2033
Qy 661 GlyAlaValLeuAlaLeuPheLeuLeuValLeuLeuValLeuValArgLysLys 680
Db 2034 GGGGCTGTCTGCTGTCTGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2093
Qy 681 ArgLysIleLysGluProLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
Db 2094 CGGAAGATCAGAGAGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAC 2153

Qy 701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db 2154 TATGGCGAAGAGGGGGTGGCGAAGAGGACCACTATGACATCACCAGCTCCACCGA 2213
Qy 721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db 2214 GGTCTGGAGGCCAGCGCGAGGTGTCTCCGCAATGAGTGGCACCAACCATCATCCG 2273
Qy 741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db 2274 ACACCCATGTACCGCTCTAGGCCAGCAACCCAGATGAATCGCAACTTTATAATTGAG 2333
Qy 761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProTyrAspThrLeuLeuValPhe 780
Db 2334 AACCTGAAGCGGCTAAACACAGACCCCGCCCGCTAGCACACCTCTTGGTGTTC 2393
Qy 781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db 2394 GACTATGAGGGCAGCGGTCCGACCGCGTCCCTGAGTCCCTCACCTCTCCCGCTCC 2453
Qy 801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheIysLysLeuAla 820
Db 2454 GACCAAGACCAAGATTAGATTATCTGAACGAGTGGGCGCCGCTTCAAGAAGCTGGCA 2513
Qy 821 AspMetTyrGlyGlyGlyGluAspAsp 829
Db 2514 GACATGTACGGTGGCGGGAGGAGCAG 2540

RESULT 12

US-10-295-027-895
; Sequence 895, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 895

354 ATCTTACGAGACACAAGAGATTGGGTGGTCTCCCAATATCTGTCTCCCTGAAATGGC 413
121 LysGlyProPheProGlnArgLeuAsnGlnLeuLysSerAsnLysAspArgAspThrLys 140
414 AAGGGTCCCTTCCCCAGAGACTGAATCAGCTCAAGTCTAATAAGATAGAGACACCAAG 473
141 IlePheTyrSerIleThrGlyProGlyAlaAspSerProGluGluGlyValPheAlaVal 160
474 ATTTTCTACAGCAATCAGGGGCCGGGGCAGACAGCCCCCTGAGGGTGTCTTCGTGTA 533
161 GluLysGluThrGlyTyrLeuLeuLeuAsnLysProLeuAspArgGluGluIleAlaLys 180
534 GAGAAGAGAGACAGCTGTGTGTGTAATAAGCCACTGGACCGGAGGAGATTGCCAAG 593
181 TyrGluLeuPheGlyHisAlaValSerGluAsnGlyAlaSerValGluAspProMetAsn 200
594 TATGAGCTCTTTGCCACAGCTGTGTGACAGAAATGGTCCCTCAGTGGAGGCCCATGAAC 653
201 IleSerIleLeuValThrAspGlnAsnAspHisLysProLysPheThrGlnAspThrPhe 220
654 ATCTCCATCATCGTACCGCAGCAATGACCAAGCCCAAGTTTACCCAGGACACCTTC 713
221 ArgGlySerValLeuGluGlyValLeuProGlyThrSerValMetGlnValThrAlaThr 240
714 CGAGGGAGTGTCTTAGAGGGAGTCTTACCAGGTACTTCTGTGATGACAGGTGACAGCCACA 773
241 AspGluAspAspAlaIleTyrThrTyrAsnGlyValValAlaTyrSerIleHisSerGln 260
774 GATGAGGATGATGCACTACACCTACCAATGGGGTGGTGTCTTACTCCATCCATAGCCAA 833
261 GluProLysAspProHisAspLeuMetPheThrIleHisArgSerThrGlyThrIleSer 280
834 GAACCAAGAGCCACACGACCTCATGTTCCACAAATTCACCGGAGCACAGGCACCATCAGC 893
281 ValIleSerSerGlyLeuAspArgGluLysValProGluTyrThrLeuThrIleGlnAla 300
894 GTATCTCCAGTGGCTGGACCGGGAAGATCTCTGAGTACACACTGACCATCCAGGCC 953
301 ThrAspMetAspGlyAspGlySerThrThrAlaValAlaValValGluIleLeuAsp 320
954 ACAGACATGATGGGACGGCTCCACCACCGCAGTGCAGTAGTGAGATCTTGAT 1013
321 AlaAsnAspAsnAlaProMetPheAspProGlnLysTyrGluAlaHisValProGluAsn 340
1014 GCCAATGCAATGCTCCATGTTTGACCCCAAGTACAGGCCCATGTGCTGAGAT 1073
341 AlaValGlyHisGluValGlnArgLeuThrValThrAspLeuAspAlaProAsnSerPro 360
1074 GCAGTGGCCATGAGGTGCAGAGGCTGACGGTCACTGATCTGACGCCCACTCACC 1133
361 AlaTyrArgAlaThrTyrLeuIleMetGlyGlyAspAspGlyAspHisPheThrIleThr 380
1134 GCGTGGCGTCCCACTACCTTATCATGGCGGTGACGACGGGACCATTTTACCATCACC 1193
381 ThrHisProGluSerAsnGlnGlyIleLeuThrThrArgLysGlyLeuAspPheGluAla 400
1194 ACCACCTTGAGACCAACGAGGCACTCTGACAAACGAGAGGGTTTGATTTGAGGCC 1253
401 LysAsnGlnHisThrLeuTyrValGluValThrAsnGluAlaProPheValLeuLysLeu 420
1254 AAAAAACGACACACCTGTACGTTGAAGTACCAACGAGGCCCTTTGTGCTGAAGCTC 1313
421 ProThrSerThrAlaThrIleValHisValGluAspValAsnGluAlaProValPhe 440
1314 CCAACTCCACAGCCACATAGTGGTCCAGAGGGCATCCCCACTGGGAGGCTGTGTGT 1373
441 ValProSerLysValValGluValGlnGluGlyIleProThrGlyGluProValCys 460
1374 GTCCCACTCCAAAGTGTGTGAGTCCAGAGGGCATCCCCACTGGGAGGCTGTGTGT 1433
461 ValTyrThrAlaGluAspProAspLysGluAsnGlnLysIleSerTyrArgIleLeuArg 480
1434 GTCTACACTGCAGAACCCCTGACAAAGGAGAAATCAAAAGATACGTACCGCATCTCTGAGA 1493

481 AspProAlaGlyTyrIleLeuAlaMetAspProAspSerGlyGlnValThrAlaValGlyThr 500
1494 GACCCAGCAGGGTGGCTAGCCATGGACCCAGACAGTGGSCAGGTACAGCTGTGGCACC 1553
501 LeuAspArgGluAspGluGlnPheValArgAsnAsnIleTyrGluValMetValLeuAla 520
1554 CTCGACCGTGCAGATGAGCAGTGTGTGAGGAACAACATCTATGAAGTCAATGGCTTGGCC 1613
521 MetAspAsnGlySerProProThrThrGlyThrGlyThrLeuLeuLeuThrLeuLeuAsp 540
1614 ATGGCAATGGAAGCCCTCCACCACTGGCAACGGAAACCTTCTCTCTAACAACATGATTGAT 1673
541 ValAsnAspHisGlyProValProGluProArgGlnIleThrIleCysAsnGlnSerPro 560
1674 GTCAACGACCATGGCCCGAGTCCCTGAGCCCGTCAGATCACCATCTGCAACCAAGCCCT 1733
561 ValArgHisValLeuAsnIleThrAspLysAspLeuSerProHisThrSerProPheGln 580
1734 GTGGCCACGCTGCTGAACATCATCGGACAGGACCTGTCTCCCAACACCTCCCTTCCAG 1793
581 AlaGlnLeuThrAspAspSerAspIleTyrThrAlaGluValAsnGluGlyAsp 600
1794 GCCCAGCTCACAGATGACTCAGACATCTACTGACGGCAGAGGTCAACGAGGAAGGTGAC 1853
601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
1854 ACAGTGGTCTTGTCCCTGAAGAAGTCTCTGAAGCAGGATACATATGACGTGCACCTTCT 1913
621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
1914 CTGTCTGACCATGGCAACAAGAGCAGCTGACGGTGTATCAGGGCCACTGTGTGCACATGC 1973
641 HisGlyHisValGluThrCysProGlyProTyrLysGlyGlyPheIleLeuProValLeu 660
1974 CATGGCCATGTGCAAACTGCCCTGGACCTGGAAAGAGGTTTCTCTCTCTCTGTGTG 2033
661 GlyAlaValLeuAlaLeuPheLeuLeuValLeuLeuLeuValLeuValArgLysLys 680
2034 GGGGCTGTCTGGCTCTGTCT 2093
681 ArgLysIleLysGluProLeuLeuLeuLeuProGluAspAspThrArgAspAsnValPheTyr 700
2094 CGGAAGATCAAGAGGCCCTCTCTCTCCAGAAGATGACCCCGTGACACCGTCTCTCTCTAC 2153
701 TyrGlyGluGluGlyGlyGlyGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
2154 TATGGCGAAGAGGGGGTGGCGAAGAGGACCCAGGACTATGACATCACCAGCTCCACCGA 2213
721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
2214 GGTCTGGAGGCCAGGCCGAGGTGTCTCTCGCAATGACGTGGCACCACCAACCATCATCCCG 2273
741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
2274 ACACCATGTACCGTCTTAGCCGACCAACCCAGATGAAATCGGCACTTTATTAATTGAG 2333
761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
2334 AACCTGAAGGGGTAAACACAGACCCCAAGCCCGCCCTACGACACCTCTTGTGTGTTTC 2393
781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
2394 GACTATAGGGGAGCGGCTCCGACCGCGCTCCCTGAGCTCCCTCACCCTCTCTCGCCCTCC 2453
801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTyrGlySerArgPheLysLeuAla 820
2454 GACCAAGACCAAGATTACGATTATCTGAACGAGTGGGGCAGCCCGCTTCAAGAGCTGGCA 2513
821 AspMetTyrGlyGlyGluAspAsp 829
2514 GACATGTACGTGGGGGAGGACGAC 2540


```
Db      ||||| GTGGCCACGTGTGTAACATCAGGACAGGACCTGTCTCCCCACACCTCCCTTTCCAG 2355
QY      581 AlaGlnLeuThrAspAspSerAspIleTyrTrpThrAlaGluValAsnGluGlyAsp 600
Db      2356 GCCAGCTCACAGATGACTCAGACATCTACTGGACGCGAGGTCAACGAGGAAGGTGAC 2415
QY      601 ThrValValLeuSerLeuLysLysPheLeuLysGlnAspThrTyrAspValHisLeuSer 620
Db      2416 ACAGTGTCTTGTCCCTGAAGAGTTCCTGAGCAGGATACATATGACGTGCACCTTTCT 2475
QY      621 LeuSerAspHisGlyAsnLysGluGlnLeuThrValIleArgAlaThrValCysAspCys 640
Db      2476 CTGTCTGACCATGCACAAGAGCAGCTGACGGTGCATCAGGGCCACTGTGTGCGACTGC 2535
QY      641 HisGlyHisValGluThrCysProGlyProTyrPheGlyGlyPheIleLeuProValLeu 660
Db      2536 CATGGCCATGTCGAAACCTCCCTGGACCTCGAAGGAGGTTTCATCCCTCCCTGTGCTG 2595
QY      661 GlyAlaValLeuAlaLeuPheLeuLeuValLeuLeuLeuValArgLysLys 680
Db      2596 GGGGCTGTCTGTGCTGTCTGCTCTCTCTGCTGCTGCTTTTGTGGTGAGAAAGAG 2655
QY      681 ArgLysIleLysGluProLeuLeuLeuProGluAspThrArgAspAsnValPheTyr 700
Db      2656 CGGAAGATCAAGGAGGCCCCCTCTACTCCCAAGAGATGACACCCGTGACACCGTCTCTAC 2715
QY      701 TyrGlyGluGluGlyGlyGluGluAspGlnAspTyrAspIleThrGlnLeuHisArg 720
Db      2716 TATGGCGAAGAGGGGTGGCGAAGAGGACGAGGACTATGACATCACCCAGCTCCACCGA 2775
QY      721 GlyLeuGluAlaArgProGluValValLeuArgAsnAspValAlaProThrIleIlePro 740
Db      2776 GGTCGTGAGGCCAGCGCGGAGGTGGTCTCCGCAATGACGTGGCACCAACCATCATCCCG 2835
QY      741 ThrProMetTyrArgProArgProAlaAsnProAspGluIleGlyAsnPheIleIleGlu 760
Db      2836 ACACCCATGTACCGTCTAGCCGACGCCAACCCAGATGAAATCGGCACCTTTATATTGAG 2895
QY      761 AsnLeuLysAlaAlaAsnThrAspProThrAlaProProTyrAspThrLeuLeuValPhe 780
Db      2896 AACCTGAAGCGCGCTAACACAGACCCGACAGCCCGCCCTACGACACCCCTCTGGTGTTC 2955
QY      781 AspTyrGluGlySerGlySerAspAlaAlaSerLeuSerSerLeuThrSerSerAlaSer 800
Db      2956 GACTATGAGGGCAGCGCTCCGACGCGCGCTCCCTGAGCTCCCTCCTCCCTCCCGCTCC 3015
QY      801 AspGlnAspGlnAspTyrAspTyrLeuAsnGluTrpGlySerArgPheLysLysLeuAla 820
Db      3016 GACCAAGACCAAGATTACGATTATCTGACGAGTGGGGCAGCGCTTCAAGAAGCTGGCA 3075
QY      821 AspMetTyrGlyGlyGluAspAsp 829
Db      3076 GACATGTACGGTGGCGGAGGAGCAGAC 3102
```

Search completed: September 23, 2004, 10:29:38
Job time : 1895 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:18:25 ; Search time 46 Seconds
(without alignments)
1733.539 Million cell updates/sec

Title: US-09-916-849A-1
Perfect score: 829
Sequence: 1 MGLPRGLASLLLLQVCWLQ.....NEWGSRFKLADWYGGGDD 829

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 78:**
1: Pir1:**
2: Pir2:**
3: Pir3:**
4: Pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	829	100.0	829	1 IJHUCP	cadherin 3 precurs
2	61	7.4	491	1 IJBOCP	P-cadherin - bovin
3	51	6.2	822	1 IJMSCP	P-cadherin precurs
4	17	2.1	732	1 IJCHCB	B-cadherin precurs
5	17	2.1	887	1 IJCHCL	E-cadherin precurs
6	16	1.9	895	1 IJXUCP	Ep-cadherin precurs
7	16	1.9	905	2 S43064	cadherin - African
8	15	1.8	871	2 S47518	cadherin - African
9	15	1.8	882	1 IJHUCE	cadherin 1 precurs
10	15	1.8	884	1 IJMSCE	E-cadherin precurs
11	15	1.8	894	2 S34438	uvomorulin - mouse
12	13	1.6	913	1 A75433	R-cadherin precurs
13	13	1.6	913	1 IJCHCR	R-cadherin precurs
14	13	1.6	916	2 C38992	cadherin 4 precurs
15	12	1.4	783	2 I50116	N-cadherin precurs
16	12	1.4	877	1 IJBOCN	N-cadherin precurs
17	12	1.4	905	1 IJXLC1	N-cadherin 1 precu
18	12	1.4	906	1 IJHUCN	cadherin 2 precurs
19	12	1.4	906	1 IJWSCN	N-cadherin precurs
20	12	1.4	906	1 IJXLC2	N-cadherin 2 precu
21	12	1.4	912	1 IJCHCN	N-cadherin precurs
22	11	1.3	789	2 I52701	K-cadherin - rat
23	11	1.3	790	2 I37016	cadherin-6 - human
24	10	1.2	353	2 T07998	cytochrome c-type
25	10	1.2	353	2 T08001	cytochrome c-type
26	10	1.2	790	2 I50178	cadherin-6B - chic
27	9	1.1	481	2 C95920	hypothetical membr
28	9	1.1	3034	2 T14119	seven-pass transme
29	8	1.0	40	2 D83378	hypothetical prote

30 8 1.0 103 2 A26736 transformation-ind
31 8 1.0 132 2 H75548 hypothetical prote
32 8 1.0 139 2 F85565 protein R10E11.9 [
33 8 1.0 157 2 A86413 probable ribosomal
34 8 1.0 182 2 T46396 hypothetical prote
35 8 1.0 243 2 D64188 arginine transport
36 8 1.0 279 2 G83196 glycerol uptake fa
37 8 1.0 293 2 T32892 hypothetical prote
38 8 1.0 340 2 T32891 hypothetical prote
39 8 1.0 364 2 F70137 cell division prot
40 8 1.0 379 2 T11375 ubiquinol-cytochro
41 8 1.0 384 2 B83443 probable acyl-CoA
42 8 1.0 393 2 T01371 hypothetical prote
43 8 1.0 398 2 S77230 hypothetical prote
44 8 1.0 463 2 AF0469 probable permease
45 8 1.0 518 2 S55948 hypothetical prote

ALIGNMENTS

RESULT 1

IJHUCP
cadherin 3 precursor - human
N:Alternate names: P-cadherin; placental cadherin
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C:Accession: A33659
R:Shimoyama, Y.; Yoshida, T.; Terada, M.; Shimozato, Y.; Abe, O.; Hirohashi, S.
J. Cell Biol. 109, 1787-1794, 1989
A:Title: Molecular cloning of a human Ca(2+)-dependent cell-cell adhesion molecule homolo
A:Reference number: A33659; MUID:90009051; PMID:2793940
A:Accession: A33659
A:Molecule type: mRNA
A:Residues: 1-829 <SHI>
A:Cross-references: GB:X63629; NID:G35322; PIDN:CAA45177.1; PID:G35323
C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to
C:Genetics:
A:Gene: GDB:CDH3
A:Cross-references: GDB:I32860; OMIM:114021
A:Map position: 16q24.1-16qter
C:Superfamily: cadherin; cadherin repeat homology
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane prot
F:1-46/Domain: signal sequence #status predicted <SIG>
F:127-107/Domain: propeptide #status predicted <PRO>
F:108-829/Product: P-cadherin #status predicted <MAT>
F:110-215/Domain: extracellular #status predicted <EXT>
F:185-190/Region: cadherin repeat homology <CR1>
F:218-328/Domain: cadherin binding #status predicted
F:331-440/Domain: cadherin repeat homology <CR2>
F:441-548/Domain: cadherin repeat homology <CR3>
F:549-652/Domain: cadherin repeat homology <CR4>
F:653-677/Domain: transmembrane #status predicted <TMM>
F:678-829/Domain: intracellular #status predicted <INT>
F:785-800/Region: serine-rich
F:200,566/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 829; DB 1; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGLPRGLASLLLLQVCWLQCAASEPCRAVFRFAEVLTEAGGAQEPQALGKVMGCPG 60
Db 1 MGLPRGLASLLLLQVCWLQCAASEPCRAVFRFAEVLTEAGGAQEPQALGKVMGCPG 60
Qy 61 QEPALFSTDNDFTVRNGETVQERRSLKRNPLKIFPSKILRHRKRDVWVAPISVPENG 120
Db 61 QEPALFSTDNDFTVRNGETVQERRSLKRNPLKIFPSKILRHRKRDVWVAPISVPENG 120
Qy 121 KGPFPQRLNQLKSKNDRDTKIFYISITGPGADSPPEGVFAVEKETGWLILNKLPLDREIAK 180
Db 121 KGPFPQRLNQLKSKNDRDTKIFYISITGPGADSPPEGVFAVEKETGWLILNKLPLDREIAK 180

Query Match 7.4%; Score 61; DB 1; Length 491;
Best Local Similarity 100.0%; Pred. No. 1.3e-51;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 YELFGHVAENGASVEDPNNISIIIVTDQNDHKKPKFTQDTRFRGSVLEGVLPGTSMQVNTAT 240
DB 181 YELFGHVAENGASVEDPNNISIIIVTDQNDHKKPKFTQDTRFRGSVLEGVLPGTSMQVNTAT 240

QY 241 DEDALITYNGVAVYSHSQEPKDPHLMFTTHRSSTGTSVSSGLDREKVPYETLTQA 300
DB 241 DEDALITYNGVAVYSHSQEPKDPHLMFTTHRSSTGTSVSSGLDREKVPYETLTQA 300

QY 301 TMDGSGSTTTAVAVVEILLDANDAMPDPQKYEAHVAVENAVGHEVQRLTVTDLPNSP 360
DB 301 TMDGSGSTTTAVAVVEILLDANDAMPDPQKYEAHVAVENAVGHEVQRLTVTDLPNSP 360

QY 361 AWRATVLMGDDGDHFTTITTHPESNOGILTTTRKGLDFAKQHTLYVEVINEAPFVLKL 420
DB 361 AWRATVLMGDDGDHFTTITTHPESNOGILTTTRKGLDFAKQHTLYVEVINEAPFVLKL 420

QY 421 PTSTATVHVHVEDVNEAPVFPVPSKVEVQEGIPTGEPVCVYTAEDPDKENOKISYRIILR 480
DB 421 PTSTATVHVHVEDVNEAPVFPVPSKVEVQEGIPTGEPVCVYTAEDPDKENOKISYRIILR 480

QY 481 DPAGWLAMDPSQCVTAAGTLDREDFQVRNNIYVWMLAMONGSPPTGTGLLTLLD 540
DB 481 DPAGWLAMDPSQCVTAAGTLDREDFQVRNNIYVWMLAMONGSPPTGTGLLTLLD 540

QY 541 VNDHGPVPEPQITICNOSPVRHLNITDKLSPHTSFQAQLTDSDIYWTAEVNEEGD 600
DB 541 VNDHGPVPEPQITICNOSPVRHLNITDKLSPHTSFQAQLTDSDIYWTAEVNEEGD 600

QY 601 TVVLSLKKFLKQDYVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGWKGGFLPVL 660
DB 601 TVVLSLKKFLKQDYVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGWKGGFLPVL 660

QY 661 GAVLALLFLLVLLVLRKKRKKKEPLLIPEDDTRDNVFFYGGEGGEDQDYDITQLHR 720
DB 661 GAVLALLFLLVLLVLRKKRKKKEPLLIPEDDTRDNVFFYGGEGGEDQDYDITQLHR 720

QY 721 GLEAPPEVLRNDVAPTIPTMYRPRANPDPEIGNFIENKAANTDPTAPPYDTLLVF 780
DB 721 GLEAPPEVLRNDVAPTIPTMYRPRANPDPEIGNFIENKAANTDPTAPPYDTLLVF 780

QY 781 DYEGSGDAASLSLTSSASDQDDYDLNWSGRFKLADMTGGGDD 829
DB 781 DYEGSGDAASLSLTSSASDQDDYDLNWSGRFKLADMTGGGDD 829

RESULT 2
LJBOCP
P-cadherin - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C:Accession: S11694
R:Liaw, C.W.; Cannon, C.; Power, M.D.; Kiboneka, P.K.; Rubin, L.L.
EMBO J. 9, 2701-2708, 1990
A:Title: Identification and cloning of two species of cadherins in bovine endothelial cells
A:Reference number: S11693; MUID:90360979; PMID:2390969
A:Accession: S11694
A:Molecule type: mRNA
A:Residues: 1-491 <Lia>
A:Cross-references: EMBL:X53614; NID:g166; PIDN:CRA37676.1; PID:g833777
C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in cell signaling.
C:Superfamily: cadherin; cadherin repeat homology
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein
F:1-491/Product: P-cadherin (fragment) #status predicted <MAT>
F:1-314/Domain: extracellular (fragment) #status predicted <CR3>
F:1-102/Domain: cadherin repeat homology (fragment) #status predicted <CR4>
F:103-210/Domain: cadherin repeat homology <CR5>
F:211-314/Domain: cadherin repeat homology <CR5>
F:315-339/Domain: transmembrane #status predicted <TMM>
F:340-431/Domain: intracellular #status predicted <INT>
F:447-462/Region: serine-rich
F:228/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 6.2%; Score 51; DB 1; Length 822;
Best Local Similarity 100.0%; Pred. No. 1.6e-41;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 684 KEPLLPEDDTRDNVFFYGGEGGEDQDYDITQLHRGLEAPPEVLRNDV 734
DB 677 KEPLLPEDDTRDNVFFYGGEGGEDQDYDITQLHRGLEAPPEVLRNDV 727

RESULT 4
LJCHCB
B-cadherin precursor - chicken (fragment)
N:Alternate names: K-CAM protein
C:Species: Gallus gallus (chicken)
C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C:Accession: A41634; A38715; S16160
R:Sorkin, B.C.; Gallin, W.J.; Edelman, G.M.; Cunningham, B.A.
Proc. Natl. Acad. Sci. U.S.A. 88, 11545-11549, 1991
A:Title: Genes for two calcium-dependent cell adhesion molecules have similar structures

Query Match 7.4%; Score 61; DB 1; Length 491;
Best Local Similarity 100.0%; Pred. No. 1.3e-51;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 676 LVRRKKRIKPEPLLPEDDTRDNVFFYGGEGGEDQDYDITQLHRGLEAPPEVLRNDV 735
DB 338 LVRRKKRIKPEPLLPEDDTRDNVFFYGGEGGEDQDYDITQLHRGLEAPPEVLRNDV 397

QY 736 P 736
DB 398 P 398

RESULT 3
LJMSCP
P-cadherin precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 07-May-1999
C:Accession: S03163; S34458
R:Nose, A.; Nagafuchi, A.; Takeichi, M.
EMBO J. 6, 3655-3661, 1987
A:Title: Isolation of placental cadherin cDNA: identification of a novel gene family of (C) type I transmembrane proteins
A:Reference number: S03163; MUID:98111554; PMID:3428270
A:Accession: S03163
A:Molecule type: mRNA
A:Residues: 1-822 <NOS>
A:Cross-references: EMBL:X06340
R:Faraldo, M.L.M.; Cano, A.
J. Mol. Biol. 231, 935-941, 1993
A:Title: The 5' flanking sequences of the mouse P-cadherin gene. Homologies to 5' sequences of other genes
A:Reference number: S34458; MUID:93294853; PMID:8515462
A:Accession: S34458
A:Molecule type: DNA
A:Residues: 1-55 <FAR>
A:Cross-references: EMBL:X68057
C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in cell signaling.

A:Reference number: A41634; MUID:92107987; PMID:1763068
 A:Accession: A41634
 A:Molecule type: DNA
 A:Residues: 1-732 <SOR>
 A:Cross-references: GB:M81894; NID:9212226; PIDN:AAA48925.1; PID:9212227
 R:Napolitano, E.W.; Venstrom, K.; Wheeler, E.F.; Reichardt, L.F.
 J. Cell Biol. 113, 893-905, 1991
 A:Title: Molecular cloning and characterization of B-cadherin, a novel chick cadherin.
 A:Reference number: A38715; MUID:91225083; PMID:2026653
 A:Accession: A38715
 A:Molecule type: mRNA
 A:Residues: 7-413, 'V', 415-732 <NAP>
 A:Cross-references: GB:X58518; NID:9631113; PIDN:CAA41408.1; PID:9631114
 C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in a variety of other cellular processes.
 C:Keywords: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in a variety of other cellular processes.
 A:Gene: K-CAM
 A:Introns: 29/3; 81/3; 130/1; 188/3; 231/3; 292/3; 375/2; 423/1; 498/1; 571/1; 614/3; 68/3
 C:Superfamily: cadherin; cadherin repeat homology
 C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein
 F:1-6/Domain: extracellular #status predicted <PRO>
 F:6-554/Domain: extracellular #status predicted <EXT>
 F:7-732/Product: B-cadherin #status predicted <MAT>
 F:9-114/Domain: cadherin repeat homology <CR1>
 F:84-89/Region: cadherin binding #status predicted
 F:117-227/Domain: cadherin repeat homology <CR2>
 F:230-339/Domain: cadherin repeat homology <CR3>
 F:340-447/Domain: cadherin repeat homology <CR4>
 F:448-552/Domain: cadherin repeat homology <CR5>
 F:555-580/Domain: transmembrane #status predicted <TM>
 F:581-732/Domain: intracellular #status predicted <INT>
 F:689-702/Region: serine-rich
 F:137,410/Binding site: carbohydrate (Asn) (covalent) #status predicted
 Query Match 2.1%; Score 17; DB 1; Length 732;
 Best Local Similarity 100.0%; Pred. No. 6e-08;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 797 SSASDQDDYDYLNEWG 813
 DB 699 SSASDQDDYDYLNEWG 715

RESULT 5
 IUXLCP
 E-cadherin precursor, hepatic - chicken
 N:Alternate names: L-CAM; liver cell adhesion molecule
 C:Species: Gallus gallus (chicken)
 C:Date: 30-Jun-1993 #sequence revision 30-Jun-1993 #text_change 25-Oct-1996
 C:Accession: A30201; A29866; E29866
 R:Sorkin, B.C.; Hemperly, J.J.; Edelman, G.M.; Cunningham, B.A.
 Proc. Natl. Acad. Sci. U.S.A. 85, 7617-7621, 1988
 A:Title: Structure of the gene for the liver cell adhesion molecule, L-CAM.
 A:Reference number: A30201; MUID:89017248; PMID:3174655
 A:Accession: A30201
 A:Molecule type: mRNA
 A:Residues: 1-81 <SOR>
 A:Cross-references: EMBL:J04074
 R:Gallin, W.J.; Sorkin, B.C.; Edelman, G.M.; Cunningham, B.A.
 Proc. Natl. Acad. Sci. U.S.A. 84, 2808-2812, 1987
 A:Title: Sequence analysis of a cDNA clone encoding the liver cell adhesion molecule, L-CAM.
 A:Reference number: A29866; MUID:87204217; PMID:3472238
 A:Accession: A29866
 A:Molecule type: mRNA
 A:Residues: 51-887 <GAL>
 A:Cross-references: EMBL:M16260
 A:Accession: B29866
 A:Molecule type: protein
 A:Residues: 161-172;323-336;386-407;533-551 <GA2>
 C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in a variety of other cellular processes.
 C:Genetics: 23/3; 62/1; 138/1 183/3; 235/3; 284/1; 342/3; 385/3; 447/3; 529/2; 577/1; 652/3
 A:Introns: 23/3; 62/1; 138/1 183/3; 235/3; 284/1; 342/3; 385/3; 447/3; 529/2; 577/1; 652/3
 C:Superfamily: cadherin; cadherin repeat homology
 C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; liver; transmembrane protein

F:1-26/Domain: signal sequence #status predicted <SIG>
 F:27-160/Domain: propeptide #status predicted <PRO>
 F:161-887/Product: E-cadherin, hepatic #status experimental <MAT>
 F:161-704/Domain: extracellular #status predicted <EXT>
 F:163-268/Domain: cadherin repeat homology <CR1>
 F:238-243/Region: cadherin binding #status predicted
 F:271-381/Domain: cadherin repeat homology <CR2>
 F:384-493/Domain: cadherin repeat homology <CR3>
 F:494-601/Domain: cadherin repeat homology <CR4>
 F:602-704/Domain: cadherin repeat homology <CR5>
 F:705-735/Domain: transmembrane #status predicted <TM>
 F:736-887/Domain: intracellular #status predicted <INT>
 F:844-857/Region: serine-rich
 F:291,346,564,643/Binding site: carbohydrate (Asn) (covalent) #status experimental
 Query Match 2.1%; Score 17; DB 1; Length 887;
 Best Local Similarity 100.0%; Pred. No. 7.1e-08;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 797 SSASDQDDYDYLNEWG 813
 DB 854 SSASDQDDYDYLNEWG 870

RESULT 6
 IUXLCP
 E-cadherin precursor - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 30-Jun-1993 #sequence revision 30-Jun-1993 #text_change 22-Jun-1999
 C:Accession: B43785; A60128
 R:Ginsberg, D.; Desimone, D.; Geiger, B.
 Development 111, 315-325, 1991
 A:Title: Expression of a novel cadherin (EP-cadherin) in unfertilized eggs and early Xenopus laevis.
 A:Reference number: A43785; MUID:91372132; PMID:1893866
 A:Accession: B43785
 A:Molecule type: mRNA
 A:Residues: 1-895 <GIN>
 A:Cross-references: GB:X63720; NID:964681; PIDN:CAA45252.1; PID:964682
 A:Note: it is uncertain whether Met-1 or Met-16 is the initiator
 R:Angres, B.; Mueller, A.H.J.; Kellermann, J.; Hausen, P.
 Development 111, 829-844, 1991
 A:Title: Differential expression of two cadherins in Xenopus laevis.
 A:Reference number: A60128; MUID:91347911; PMID:1879345
 A:Accession: A60128
 A:Molecule type: protein
 A:Residues: 171-177, 'I', 179-183, 'K', 185-189, 'XI', <ANG>
 A:Note: the material sequenced may have contained U-cadherin as well as E-cadherin
 C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to be involved in a variety of other cellular processes.
 C:Superfamily: cadherin; cadherin repeat homology
 C:Keywords: calcium binding; cell adhesion; duplication; embryo; glycoprotein; transmembrane protein
 F:1-43/Domain: signal sequence #status predicted <SIG>
 F:44-170/Domain: propeptide #status predicted <PRO>
 F:171-895/Product: EP-cadherin #status predicted <MAT>
 F:171-718/Domain: extracellular #status predicted <EXT>
 F:173-278/Domain: cadherin repeat homology <CR1>
 F:248-253/Region: cadherin binding #status predicted
 F:281-391/Domain: cadherin repeat homology <CR2>
 F:394-502/Domain: cadherin repeat homology <CR3>
 F:503-610/Domain: cadherin repeat homology <CR4>
 F:611-714/Domain: cadherin repeat homology <CR5>
 F:719-743/Domain: transmembrane #status predicted <TM>
 F:744-895/Domain: intracellular #status predicted <INT>
 F:852-865/Region: serine-rich
 F:440,696/Binding site: carbohydrate (Asn) (covalent) #status predicted
 Query Match 1.9%; Score 16; DB 1; Length 895;
 Best Local Similarity 100.0%; Pred. No. 7e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 699 FYGEGGGGEBDDYD 714
 DB 765 FYGEGGGGEBDDYD 780

RESULT 7

S43064
 Cadenherin - African clawed frog
 C;Species: Xenopus laevis (African clawed frog)
 C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
 C;Accession: S43064; S43065
 R;Kuehl, M.
 submitted to the EMBL Data Library, March 1994
 A;Reference number: S43064
 A;Accession: S43064
 A;Molecule type: mRNA
 A;Residues: 1-905 <KUE>
 A;Cross-references: EMBL:X78546; NID:G468816; PIDN:CAA55292.1; PID:G468817
 R;Herzberg, F.; Wildermuth, V.; Wedlich, D.
 Mech. Dev. 35, 33-42, 1991
 A;Title: Expression of XBCad, a novel cadherin, during oogenesis and early development
 A;Reference number: S43065; MUID:92062581; PMID:1840622
 A;Accession: S43065
 A;Molecule type: mRNA
 A;Residues: 'NSA', 462-697, 'Q', 699-807, 'A', 809-840, 'N', 842-877, 'N', 879-883, 'N', 885-902, 'E'
 A;Cross-references: EMBL:X78546
 C;Superfamily: cadherin; cadherin repeat homology
 C;Keywords: calcium binding; cell adhesion; glycoprotein; transmembrane protein
 F;181-286/Domain: cadherin repeat homology <CR1>
 F;289-399/Domain: cadherin repeat homology <CR2>
 F;402-510/Domain: cadherin repeat homology <CR3>
 F;511-618/Domain: cadherin repeat homology <CR4>
 F;619-722/Domain: cadherin repeat homology <CR5>

Query Match 1.9%; Score 16; DB 2; Length 905;
 Best Local Similarity 100.0%; Pred. No. 7.1e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 699 FYTGGGGGGEDQDYD 714
 |||||
 DB 773 FYTGGGGGGEDQDYD 788

RESULT 8

S47518
 Cadenherin - African clawed frog
 C;Species: Xenopus laevis (African clawed frog)
 C;Date: 01-Feb-1995 #sequence_revision 10-Nov-1995 #text_change 17-Mar-1999
 C;Accession: S47518
 R;Toci, O.; Fujii, G.; Tashiro, K.; Shikawa, K.
 Biochim. Biophys. Acta 1219, 121-128, 1994
 A;Title: Molecular cloning of cDNA for XTCAD-1, a novel Xenopus cadherin, and its expression
 A;Reference number: S47518; MUID:94368839; PMID:8086449
 A;Accession: S47518
 A;Molecule type: mRNA
 A;Residues: 1-871 <TOO>
 C;Superfamily: cadherin; cadherin repeat homology
 F;151-256/Domain: cadherin repeat homology <CR>

Query Match 1.8%; Score 15; DB 2; Length 871;
 Best Local Similarity 100.0%; Pred. No. 6.7e-06;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 744 YRPRPNDPEIGNFI 758
 |||||
 DB 787 YRPRPNDPEIGNFI 801

RESULT 9

LHUCE
 Cadenherin 1 precursor [validated] - human
 N;Alternate names: ARC-1; cell CAM 120/80; E-cadherin; epithelial cadherin; L-CAM; uvomor-
 C;Species: Homo sapiens (man)
 C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 08-Dec-2000
 C;Accession: S37654; S31430; S05475; S31460; S06716; A57171; JC2230; I52294; I52704; S25
 R;Bussemakers, M.J.G.; van Bokhoven, A.; Mees, S.G.M.; Kemler, R.; Schalken, J.A.
 Mol. Biol. Rep. 17, 123-128, 1993

A;Title: Molecular cloning and characterization of the human E-cadherin cDNA.
 A;Reference number: S37654; MUID:93211394; PMID:8459805
 A;Accession: S37654
 A;Molecule type: mRNA
 A;Residues: 1-882 <BUS>
 A;Cross-references: EMBL:Z13009; NID:G31072; PIDN:CAA78353.1; PID:G31073
 R;Keller, W.; Warda, A.; Oda, T.; Hirohashi, S.; Kemler, R.; Birchmeier, W.
 submitted to the EMBL Data Library, December 1992
 A;Description: Sequence of human E-cadherin cDNA.
 A;Reference number: S31430
 A;Accession: S31430
 A;Molecule type: mRNA
 A;Residues: 1-542, 'F', 544-882 <KEL>
 A;Cross-references: EMBL:Z18923; NID:G31074; PIDN:CAA79356.1; PID:G31075
 R;Mansouri, A.; Spurt, N.; Goodfellow, P.N.; Kemler, R.
 Differentiation 38, 67-71, 1988
 A;Title: Characterization and chromosomal localization of the gene encoding the human cel
 A;Reference number: S05475; MUID:89031725; PMID:3263290
 A;Accession: S05475
 A;Molecule type: mRNA
 A;Residues: 157-311 <MAN>
 A;Cross-references: EMBL:X12790
 R;Frixen, U.H.
 submitted to the EMBL Data Library, March 1990
 A;Reference number: S31460
 A;Accession: S31460
 A;Molecule type: mRNA
 A;Residues: 265-352 <FR1>
 A;Cross-references: EMBL:X52279; NID:G28821; PIDN:CAA36522.1; PID:G28822
 R;Wheeler, M.J.; Buck, C.A.; Bechtol, K.B.; Damsky, C.H.
 J. Cell. Biochem. 34, 187-202, 1987
 A;Title: Soluble 80-kd fragment of cell-CAM 120/80 disrupts cell-cell adhesion.
 A;Reference number: S06716; MUID:87280410; PMID:3611200
 A;Accession: S06716
 A;Molecule type: protein
 A;Residues: 'XQ', 157-162, 'V', 164-179 <WHE>
 R;Berx, G.; Staes, K.; van Hengel, J.; Molemans, F.; Bussemakers, M.J.G.; van Bokhoven, J
 Genomics 26, 281-289, 1995
 A;Title: Cloning and characterization of the human invasion suppressor gene E-cadherin (C
 A;Reference number: A57171; MUID:95324920; PMID:7601454
 A;Accession: A57171
 A;Status: nucleic acid sequence not shown; not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 1-30, 32-882 <BER>
 A;Cross-references: GB:L34784
 R;Rimm, D.L.; Morrow, J.S.
 Biochem. Biophys. Res. Commun. 200, 1754-1761, 1994
 A;Title: Molecular cloning of human E-cadherin suggests a novel subdivision of the cadhe
 A;Reference number: JC2230; MUID:94242050; PMID:8185635
 A;Accession: JC2230
 A;Molecule type: mRNA
 A;Residues: 1-9, 'G', 11-15, 'RSPGSGRSPPCITRLHVGAPAPKPR', 52-67, 'I', 69, 'LTPIP', 76-94,
 A;Cross-references: GB:L08599; NID:G340184; PIDN:AAA61259.1; PID:G340185
 A;Note: the majority of differences between this and other reports represent apparent fra
 A;Note: the authors translated the codon CCG for residue 868 as Arg
 R;Bussemakers, M.J.G.; Groidi, L.A.; van Bokhoven, A.; Schalken, J.A.
 Biochem. Biophys. Res. Commun. 203, 1284-1290, 1994
 A;Title: Transcriptional regulation of the human E-cadherin gene in human prostate cancer
 A;Reference number: I52294; MUID:94380041; PMID:8093045
 A;Accession: I52294
 A;Status: translation not shown; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-16 <RES>
 A;Cross-references: GB:L34545; NID:G509604; PIDN:AAA21764.1; PID:G509605
 R;Becker, K.F.; Atkinson, M.J.; Reich, U.; Nekarda, H.; Stewert, J.R.; Hofler
 Cancer Res. 54, 3845-3852, 1994
 A;Title: E-cadherin gene mutations provide clues to diffuse type gastric carcinomas.
 A;Reference number: I52704; MUID:94306394; PMID:8033105
 A;Accession: I52704
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 337-476 <RE2>

A;Cross-references: GB:S72492; NID:G632756; PIDN:AAI4108.1; PID:G4261808
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion and are thought to
C;Genetics:
A;Gene: GDB:CDH1; UVO
A;Cross-references: GDB:120484; OMIM:192090
A;Map position: 16q22.1-16q22.1
A;Introns: 379/3; 440/3
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein
F;1-27/Domain: signal sequence #status predicted <SIG>
F;128-156/Domain: propeptide #status predicted <PRO>
F;157-699/Domain: extracellular #status predicted <CR1>
F;157-699/Product: E-cadherin; epithelial #status experimental <EAT>
F;267-377/Domain: cadherin repeat homology <CR2>
F;380-488/Domain: cadherin repeat homology <CR3>
F;489-597/Domain: cadherin repeat homology <CR4>
F;598-702/Domain: cadherin repeat homology <CR5>
F;702-733/Domain: transmembrane #status predicted <TM>
F;734-884/Domain: intracellular #status predicted <INT>
F;842-855/Region: serine-rich
F;840-853/Region: serine-rich
F;376,558,570,622,637,849/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 1.8%; Score 15; DB 1; Length 882;
Best Local Similarity 100.0%; Pred. No. 6.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 RPKLADMYGGEDD 829
DB 868 RPKLADMYGGEDD 882

RESULT 10
IUMSCE
E-cadherin precursor, epithelial - mouse
N;Alternate names: uvomorulin
C;Species: Mus musculus (house mouse)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: S04528; S03160; I49565; S48735
R;Nagafuchi, A.; Shirayoshi, Y.; Okazaki, M.
Nature 329, 341-343, 1987
A;Title: Transformation of cell adhesion properties by exogenously introduced E-cadherin
A;Reference number: S04528; MUID:87315445; PMID:3498123
A;Molecule type: mRNA
A;Accession: S04528
A;Residues: 1-412, 'V', 414-884 <NAG>
A;Cross-references: EMBL:X06115
R;Ringwald, M.; Schuh, R.; Vestweber, D.; Bistetter, H.; Lottspeich, F.; Engel, J.; Doel
EMBO J. 6, 3647-3653, 1987
A;Title: The structure of cell adhesion molecule uvomorulin. Insights into the molecular
A;Reference number: S03160; MUID:8811553; PMID:3501370
A;Accession: S03160
A;Molecule type: mRNA
A;Residues: 157-884 <RIN>
A;Cross-references: EMBL:X06339
A;Note: part of this sequence, including the amino end of the mature protein, was confir
R;Behrens, J.; Loewrick, O.; Klein-Hitpass, L.; Birchmeier, W.
Proc. Natl. Acad. Sci. U.S.A. 88, 11495-11499, 1991
A;Title: The E-cadherin promoter: Functional analysis of a G-C-rich region and an epithe
A;Reference number: I49565; MUID:92107977; PMID:1763063
A;Accession: I49565
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-15 <RES>
A;Cross-references: GB:M81449; NID:G192325; PIDN:AAA37352.1; PID:G192326
R;Tong, K.I.; Yau, P.; Overduin, M.; Bagby, S.; Porumb, T.; Takeichi, M.; Ikura, M.
FEBS Lett. 352, 318-322, 1994
A;Title: Purification and spectroscopic characterization of a recombinant amino-terminal
A;Reference number: S48735; MUID:95010732; PMID:7925993
A;Accession: S48735
A;Status: preliminary
A;Molecule type: protein

A;Residues: 156-300 <TON>
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to
C;Genetics:
A;Gene: E-cadherin
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein
F;1-27/Domain: signal sequence #status predicted <SIG>
F;128-156/Domain: propeptide #status predicted <PRO>
F;157-699/Domain: extracellular #status predicted <CR1>
F;157-699/Product: E-cadherin; epithelial #status experimental <EAT>
F;267-377/Domain: cadherin repeat homology <CR2>
F;380-488/Domain: cadherin repeat homology <CR3>
F;489-597/Domain: cadherin repeat homology <CR4>
F;598-702/Domain: cadherin repeat homology <CR5>
F;702-733/Domain: transmembrane #status predicted <TM>
F;734-884/Domain: intracellular #status predicted <INT>
F;842-855/Region: serine-rich
F;840-853/Region: serine-rich
F;560,639/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 1.8%; Score 15; DB 1; Length 884;
Best Local Similarity 100.0%; Pred. No. 6.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 744 YRPRPNDPEIGNFI 758
DB 799 YRPRPNDPEIGNFI 813

RESULT 11
S34438
uvomorulin - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Nov-1993 #sequence_revision 03-Nov-1995 #text_change 23-May-1997
C;Accession: S34438
R;Ringwald, M.; Baribault, H.; Schmidt, C.; Kemler, R.
Nucleic Acids Res. 19, 6533-6539, 1991
A;Title: The structure of the gene coding for the mouse cell adhesion molecule uvomorulin
A;Reference number: S34438; MUID:92093614; PMID:1754391
A;Accession: S34438
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-884 <RIN>
A;Cross-references: EMBL:X60975
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1992
C;Genetics:
A;Introns: 16/3; 57/1; 131/3; 179/3; 231/3; 280/1; 338/3; 381/3; 442/3; 524/2; 573/1; 64
C;Superfamily: cadherin; cadherin repeat homology
F;380-488/Domain: cadherin repeat homology <CR3>

Query Match 1.8%; Score 15; DB 2; Length 884;
Best Local Similarity 100.0%; Pred. No. 6.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 744 YRPRPNDPEIGNFI 758
DB 799 YRPRPNDPEIGNFI 813

RESULT 12
A47543
R-cadherin precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 03-May-1994 #sequence_revision 26-May-1994 #text_change 22-Jun-1999
C;Accession: A47543; I55501
R;Hutton, J.C.; Christofori, G.; Chi, W.Y.; Edman, U.; Guest, P.C.; Hanahan, D.; Kelly, F.
Mol. Endocrinol. 7, 1151-1160, 1993
A;Title: Molecular cloning of mouse pancreatic islet R-cadherin: differential expression
A;Reference number: A47543; MUID:94067164; PMID:8247017
A;Accession: A47543
A;Molecule type: mRNA
A;Residues: 1-913 <HUT>

A; Cross-references: EMBL:G69966; NID:G429111; PIDN:CRAA9589.1; PID:G429112
R.; Matsunami, H.; Miyatani, S.; Inoue, T.; Copeland, N.; Gilbert, D.; Jenkins, N.; Takeichi,
J. Cell Sci. 106, 401-409, 1993

A; Title: Cell binding specificity of mouse R-cadherin and chromosomal mapping of the gene
A; Reference number: I55501; MUID:94095672; PMID:8270638

A; Accession: I55501

A; Status: preliminary; translated from GB/EMBL/DDBB

A; Molecule type: mRNA

A; Residues: 1-913 <RES>

A; Cross-references: GB:I41888; NID:G457658; PIDN:BAA03605.1; PID:G457659

C; Comment: Cadherins mediate calcium-dependent intercellular adhesion and are thought to
C; Superfamily: cadherin; cadherin repeat homology

C; Keywords: calcium binding, cell adhesion; duplication; glycoprotein; transmembrane pro
F; 1-27/Domain: signal sequence #status predicted <SIG>
F; 28-166/Domain: propeptide #status predicted <PRO>
F; 167-913/Product: R-cadherin #status predicted <MAT>
F; 167-721/Domain: extracellular #status predicted <EXT>
F; 169-274/Domain: cadherin repeat homology <CR1>
F; 169-274/Region: cadherin binding #status predicted
F; 244-249/Region: cadherin repeat homology <CR2>
F; 277-389/Domain: cadherin repeat homology <CR3>
F; 300-304/Domain: calcium binding #status predicted <CAB>
F; 332-504/Domain: cadherin repeat homology <CR3>
F; 507-612/Domain: cadherin repeat homology <CR4>
F; 613-721/Domain: cadherin repeat homology <CR5>
F; 722-753/Domain: transmembrane #status predicted <TM>
F; 754-913/Domain: intracellular #status predicted <INT>
F; 870-885/Region: serine-rich
F; 280, 409, 554, 629, 658, 699/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 1.6%; Score 13; DB 1; Length 913;
Best Local Similarity 100.0%; Pred. No. 0.0066;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 815 RFKKLDMMYGGGE 827
| | | | |
Db 899 RFKKLDMMYGGGE 911

RESULT 13

LUCR

R-cadherin precursor - chicken

C; Species: Gallus gallus (chicken)

C; Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 16-Jun-2000

C; Accession: JH0424

R; Imizuka, H.; Miyatani, S.; Takeichi, M.
Neuron 7, 69-79, 1991

A; Title: R-cadherin: a novel Ca2+-dependent cell-cell adhesion molecule expressed in the
A; Reference number: JH0424; MUID:91299341; PMID:1712604

A; Accession: JH0424

A; Molecule type: mRNA

A; Residues: 1-913 <INU>

A; Cross-references: GB:D14459; GB:D00849; NID:G222854; PIDN:BAA03356.1; PID:G222855

A; Experimental source: retina

C; Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought b
C; Superfamily: cadherin; cadherin repeat homology

C; Keywords: calcium binding; cell adhesion; duplication; glycoprotein; retinal; transmembr
F; 1-26/Domain: signal sequence #status predicted <SIG>
F; 27-166/Domain: propeptide #status predicted <PRO>
F; 167-913/Product: R-cadherin #status predicted <MAT>
F; 167-721/Domain: extracellular #status predicted <EXT>
F; 169-274/Domain: cadherin repeat homology <CR1>
F; 244-249/Region: cadherin binding #status predicted
F; 277-389/Domain: cadherin repeat homology <CR2>
F; 332-504/Domain: cadherin repeat homology <CR3>
F; 507-612/Domain: cadherin repeat homology <CR4>
F; 613-721/Domain: cadherin repeat homology <CR5>
F; 722-753/Domain: transmembrane #status predicted <TM>
F; 754-913/Domain: intracellular #status predicted <INT>
F; 870-885/Region: serine-rich
F; 280, 409, 554, 629, 658, 699/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 1.6%; Score 13; DB 1; Length 913;
Best Local Similarity 100.0%; Pred. No. 0.0066;

Best Local Similarity 100.0%; Pred. No. 0.0056;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 703 EGGGSEDDDYD 714
Db 652 EGGGSEDDDYD 663

Search completed: September 21, 2004, 22:26:48
Job time : 47 secs.

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:17:15 ; Search time 26 Seconds
(without alignments)

1660.237 Million cell updates/sec

Title: US-09-916-849A-1

Perfect score: 829

Sequence: 1 MGIPRCPLASILLQVCWLQ.....NEWGRFXKLADMYGGREDD 829

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	829	100.0	829	1	CAD3_HUMAN	P22223 homo sapien
2	61	7.4	491	1	CAD3_BOVIN	P19535 bos taurus
3	51	6.2	822	1	CAD3_MOUSE	RC O18927 mus musculus
4	27	3.3	145	1	CAD3_PIG	O18926 sus scrofa
5	17	2.1	732	1	CAD1_CHICK	P33145 gallus gall
6	17	2.1	887	1	CAD1_CHICK	P08641 gallus gall
7	16	1.9	880	1	CAD6_XENLA	P33148 xenopus lae
8	16	1.9	884	1	CAD6_XENLA	P33152 xenopus lae
9	15	1.8	872	1	CAD1_HUMAN	P30944 xenopus lae
10	15	1.8	882	1	CAD1_HUMAN	P12830 homo sapien
11	15	1.8	884	1	CAD1_MOUSE	P09803 mus musculus
12	15	1.8	886	1	CAD4_RAT	Q9r0t4 rattus norv
13	13	1.6	913	1	CAD4_CHICK	P24503 gallus gall
14	13	1.6	913	1	CAD4_MOUSE	P39038 mus musculus
15	13	1.6	916	1	CAD4_HUMAN	P55283 homo sapien
16	12	1.4	877	1	CAD2_BOVIN	P19534 bos taurus
17	12	1.4	893	1	CAD2_BEARE	Q90275 brachydanio
18	12	1.4	905	1	CADN_XENLA	P20310 xenopus lae
19	12	1.4	906	1	CAD2_HUMAN	P19022 homo sapien
20	12	1.4	906	1	CAD2_MOUSE	P15116 mus musculus
21	12	1.4	906	1	CAD2_RAT	Q921y3 rattus norv
22	12	1.4	906	1	CAD2_XENLA	P33147 xenopus lae
23	12	1.4	912	1	CAD2_CHICK	P10288 gallus gall
24	11	1.3	789	1	CAD6_RAT	P55280 rattus norv
25	11	1.3	789	1	CAD9_HUMAN	Q9ulb4 homo sapien
26	11	1.3	790	1	CAD6_HUMAN	P55285 homo sapien
27	10	1.2	353	1	CCSA_CHLRE	P48269 chlamydomon
28	10	1.2	790	1	CAD6_CHICK	Q90762 gallus gall
29	9	1.1	381	1	CYB_DASBY	Q34300 dasyuroides
30	9	1.1	381	1	CYB_DASCR	Q34302 dasyuroides
31	9	1.1	413	1	ACD8_MOUSE	Q9d7b6 mus musculus
32	9	1.1	415	1	ACD8_HUMAN	Q9uku7 homo sapien
33	9	1.1	788	1	CADA_HUMAN	Q9y6n8 homo sapien

34	9	1.1	790	1	CAD6_MOUSE	P97326 mus musculus
35	9	1.1	818	1	CDB1_HUMAN	Q9y5f3 homo sapien
36	9	1.1	3014	1	CLR1_HUMAN	Q9yq66 homo sapien
37	9	1.1	3034	1	CLR1_MOUSE	O35161 mus musculus
38	8	1.0	93	1	GON2_RANCA	Q8d936 rana catesb
39	8	1.0	103	1	ENF1_CHICK	P08317 gallus gall
40	8	1.0	135	1	INL5_MOUSE	Q9wug6 mus musculus
41	8	1.0	243	1	AEFP_HAEIN	P45092 haemophilus
42	8	1.0	244	1	NXP2_MOUSE	Q61199 mus musculus
43	8	1.0	262	1	NXP2_HUMAN	O95156 homo sapien
44	8	1.0	264	1	NXP2_BOVIN	Q28145 bos taurus
45	8	1.0	265	1	CADA_MOUSE	P70408 mus musculus

ALIGNMENTS

RESULT 1
CAD3_HUMAN
ID CAD3_HUMAN STANDARD; PRT; 829 AA.
AC P22223;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cadherin-3 precursor (Placental-cadherin) (P-cadherin).
GN CDH3 OR CDHP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90003051; PubMed=2793940;
RA Shimoyama Y., Yoshida T., Terada M., Shimosato Y., Abe O.,
RA Hirohashi S.;
RT "Molecular cloning of a human Ca2+-dependent cell-cell adhesion
molecule homologous to mouse placental cadherin: its low expression
in human placental tissues.";
RL J. Cell Biol. 109:1787-1794 (1989).
[2]
RN SEQUENCE FROM N.A.
RP TISSUE=Testis;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., S.W.,
FAhey J., Heltan E., Kettman M., Madan A., Young A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Kraywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[3]
RN DISEASE.
RX MEDLINE=21470328; PubMed=11544476;
RA Sprecher E., Bergman R., Richard G., Lurie R., Shalev S., Cohen N.,
RA Petronius D., Shalata A., Ambinder Y., Leibur R., Perlman I., Cohen N.,
RA Szargel R.;
RT "Hypotrichosis with juvenile macular dystrophy is caused by a
mutation in CDH3, encoding P-cadherin.";
Nat. Genet. 29:134-136 (2001).
[4]

RP VARIANT HJMD HIS-503.
RX MEDLINE=22336734; PubMed=12445216;
RA Indelman M., Bergman R., Luxie R., Richard G., Miller B.,
RA Petronius D., Ciubutaro D., Leibur R., Sprecher E.;
RT "A missense mutation in CDH3, encoding P-cadherin, causes
RT hypotrichosis with juvenile macular dystrophy";
RL J. Invest. Dermatol. 119:1210-1213(2002).
CC -1- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DISEASE: Defects in CDH3 are the cause of hypotrichosis with
CC juvenile macular dystrophy (HJMD) [MIM:601553]. HJMD is a rare
CC autosomal recessive disorder characterized by early hair loss
CC heralding severe degenerative changes of the retinal macula and
CC culminating in blindness during the second to third decade of
CC life.
CC -1- SIMILARITY: Contains 5 cadherin domains.
CC -1- DATABASES: NIMES=Mutations of the CDH3 gene;
CC NOTB=Retina International's Scientific Newsletter;
CC WWW="http://www.retina-international.com/sci-news/cdh3mut.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X63629; CAA45177.1; -;
CC EMBL: BC041846; AAH41846.1; -;
CC F1R; A33659; IUHUCF.
CC HSP; P09803; ISUH.
CC Genew; HGNC:1762; CDH3.
CC MIM; 114021; -;
CC MIM; 601553; -;
CC GO; GO:0007155; P:cell adhesion; TAS.
CC InterPro; IPR002126; Cadherin.
CC InterPro; IPR002233; Cadherin_C_term.
CC Pfam; PF00028; cadherin; 5.
CC Pfam; PF01049; Cadherin_C_term; 1.
CC PRINTS; PR00205; CADHERIN.
CC SMART; SM00112; CA; 4.
CC PROSITE; PS00232; CADHERIN 1; 3.
CC PROSITE; PS00265; CADHERIN 2; 5.
CC Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
CC Signal; Disease mutation; Vision.
CC SIGNAL 1 24 POTENTIAL.
CC FT PROPEP 25 107
CC FT CHAIN 108 829
CC FT DOMAIN 108 654
CC FT TRANSMEM 655 677
CC FT DOMAIN 678 829
CC FT DOMAIN 108 215
CC FT DOMAIN 216 328
CC FT DOMAIN 329 440
CC FT DOMAIN 441 546
CC FT DOMAIN 547 650
CC FT DOMAIN 785 800
CC FT CARBOHYD 200 200
CC FT CARBOHYD 566 566
CC FT VARIANT 503 503
CC FT CONFLICT 237 237
CC FT CONFLICT 563 563
CC FT SEQUENCE 829 AA; 91427 MW; E503CFE5D981F1 CRC64;
CC
CC Query Match 100.0%; Score 829; DB 1; Length 829;
CC Best Local Similarity 100.0%; Pred. No. 0;
CC Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPIASLLLOVLCWLOCAASEPCRAVFAEYVLEAGGAEQBPQALGVKFGVCGPG 60
DB 1 MGLPRGPIASLLLOVLCWLOCAASEPCRAVFAEYVLEAGGAEQBPQALGVKFGVCGPG 60
QY 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKILRRHKRDWVAPISVPENG 120
DB 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKILRRHKRDWVAPISVPENG 120
QY 121 KGPPFQRLNQLKSNKDRDTKIFYSITGPGADSPGEGVFAVEKETGMLLNKPLDREBIAK 180
DB 121 KGPPFQRLNQLKSNKDRDTKIFYSITGPGADSPGEGVFAVEKETGMLLNKPLDREBIAK 180
QY 181 YELFGHVAENGASVEDPMNISIVTDNDHKPKFTODTFRGSLVLEGVLPGTSMVMTAT 240
DB 181 YELFGHVAENGASVEDPMNISIVTDNDHKPKFTODTFRGSLVLEGVLPGTSMVMTAT 240
QY 241 DEDDAIYTYNGWAWAYSIHSEPKDPHDLMTIHRSTGTISVISGLDREKVPYTLTQA 300
DB 241 DEDDAIYTYNGWAWAYSIHSEPKDPHDLMTIHRSTGTISVISGLDREKVPYTLTQA 300
QY 301 TDMGDSGTTTAVAVVEILDANDNAPMFDQKYEAHVPENAVGHEVORLTVTDLDAENSP 360
DB 301 TDMGDSGTTTAVAVVEILDANDNAPMFDQKYEAHVPENAVGHEVORLTVTDLDAENSP 360
QY 361 AWRATYLMGDDGDHFTITTHPESNOGILITTRKGLDFEAKNOHTLYVEVTNEAPFVKL 420
DB 361 AWRATYLMGDDGDHFTITTHPESNOGILITTRKGLDFEAKNOHTLYVEVTNEAPFVKL 420
QY 421 PTSTATVHVVEDVNEAPVFPBPKVVEQEGPTGEPVCVYTAEDPDKENQKISYRILR 480
DB 421 PTSTATVHVVEDVNEAPVFPBPKVVEQEGPTGEPVCVYTAEDPDKENQKISYRILR 480
QY 481 DPAGWLAMPDPSGQVAVGLTDRDEQFVRNNIYEVNVLAMDXGSPPTTGTGLTLLTLD 540
DB 481 DPAGWLAMPDPSGQVAVGLTDRDEQFVRNNIYEVNVLAMDXGSPPTTGTGLTLLTLD 540
QY 541 VNDHGPVPEPRQITICNQSPVRHVLNITDKLSPTSPFOAQLTDDSDIYMTAEVNEEGD 600
DB 541 VNDHGPVPEPRQITICNQSPVRHVLNITDKLSPTSPFOAQLTDDSDIYMTAEVNEEGD 600
QY 601 TVVLSLKKFLKQDIDYVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGFWKGFILPVL 660
DB 601 TVVLSLKKFLKQDIDYVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGFWKGFILPVL 660
QY 661 GAVLALLFLVLLVLLVLRKKRIKEPLLPEDDTRDNVYFYGGEGGEDQDYDITQLHR 720
DB 661 GAVLALLFLVLLVLLVLRKKRIKEPLLPEDDTRDNVYFYGGEGGEDQDYDITQLHR 720
QY 721 GLEARPEVLRNDVAPITPTPMYRPRPANDPDELGNFIENLKAANTDPTAPPYDTLLVF 780
DB 721 GLEARPEVLRNDVAPITPTPMYRPRPANDPDELGNFIENLKAANTDPTAPPYDTLLVF 780
QY 781 DYEGSGSDAASLSLTSASDQDQDYDYNWGSRFKKLADMYGGGDD 829
DB 781 DYEGSGSDAASLSLTSASDQDQDYDYNWGSRFKKLADMYGGGDD 829

RESULT 2
CAD3_BOVIN
ID CAD3_BOVIN STANDARD; PRT; 491 AA.
AC P19535;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cadherin-3 (Placental-cadherin) (P-cadherin) (Fragment).
GN CDH3 OR CDHP.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.

```

RX MEDLINE=90360979; PubMed=2390969;
RA Liaw C.W., Cannon C., Power M.D., Kiboneka P.K., Rubin L.L.;
RT "Identification and cloning of two species of cadherins in bovine
RL endothelial cells.";
CC
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X53614; CAA37676.1; -.
CC PIR; S11694; IJBOCP.
CC InterPro; IPR002126; Cadherin.
CC InterPro; IPR000233; Cadherin_C_term.
CC Pfam; PF00028; cadherin; 3.
CC Pfam; PF01049; Cadherin_C term; 1.
CC PRINTS; PR00205; CADHERIN.
CC SMART; SM00112; CA; 2.
CC PROSITE; PS00232; CADHERIN 1; 1.
CC PROSITE; PS00268; CADHERIN 2; 2.
CC Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
KW NON_TER
FT DOMAIN <1 316 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 317 339 POTENTIAL.
FT CYTOPLASM 340 491 CYTOPLASMIC (POTENTIAL).
FT DOMAIN <1 102 CADHERIN 3.
FT DOMAIN 103 208 CADHERIN 4.
FT DOMAIN 209 314 CADHERIN 5.
FT DOMAIN 447 462 SER-RICH.
FT CARBOHYD 228 228 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 491 AA; 54207 MW; 6BE0AB5918C4771 CRC64;

Query Match 7.4%; Score 61; DB 1; Length 491;
Best Local Similarity 100.0%; Pred. No. 7.4e-49;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 676 LVKKKIKPEILLPDDTRDNVFFYGGGGEDQDYDITQHRGLRPEVWLRNDVA 735
Db 338 LVKKKIKPEILLPDDTRDNVFFYGGGGEDQDYDITQHRGLRPEVWLRNDVA 397

QY 736 P 736
Db 398 P 398

RESULT 3
CAD3_MOUSE STANDARD; PRT; 822 AA.
AC P10287; Q61465;
AT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cadherin-3 precursor (Placental-cadherin) (P-cadherin).
GN CDH3 OR CDHP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8811554; PubMed=3428270;
RA Nose A., Nagatuchi A., Takeichi M.;
RT "Isolation of placental cadherin cDNA: identification of a novel gene

```

```

RT family of cell-cell adhesion molecules.";
RL EMBO J. 6:3655-3661(1987).
RN [2]
RP PARTIAL SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Embryo;
RX MEDLINE=91360343; PubMed=1886768;
RA Hatta M., Miyatani S., Copeland N.G., Gilbert D.J., Jenkins N.A.,
RA Takeichi M.;
RT "Genomic organization and chromosomal mapping of the mouse P-cadherin
RT gene.";
RL Nucleic Acids Res. 19:4437-4441(1991).
RN [3]
RP SEQUENCE OF 1-55 FROM N.A.
RC STRAIN=C57BL/6;
RA Hatta M., Takeichi M.;
RT "Complex cell type-specific transcriptional regulation by the promoter
RT and an intron of the mouse P-cadherin gene.";
RL Dev. Growth Differ. 36:509-519(1994).
RN [4]
RP DEVELOPMENTAL STAGE.
RC STRAIN=C57BL/6; TISSUE=Testis;
RX MEDLINE=97033837; PubMed=8879495;
RA Munro S.B., Blaschuk O.W.;
RT "A comprehensive survey of the cadherins expressed in the testes of
RT fetal, immature, and adult mice utilizing the polymerase chain
RT reaction.";
RL Biol. Reprod. 55:822-827(1996).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DEVELOPMENTAL STAGE: Expression is high in both fetal and newborn
CC testis but minimal in testis of 7-day-old animals. Not detected in
CC testis of 21-day-old or adult.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X06340; CAA29646.1; -.
CC EMBL; D12688; EBA02186.1; -.
CC PIR; S03183; IJMSCP.
CC HSP; P09803; 1SUH.
CC MGD; MGI:88356; Cdh3.
CC GO; GO:0005886; C:plasma membrane; IDA.
CC InterPro; IPR002126; Cadherin.
CC InterPro; IPR000233; Cadherin_C_term.
CC Pfam; PF00028; cadherin; 5.
CC Pfam; PF01049; Cadherin_C term; 1.
CC PRINTS; PR00205; CADHERIN.
CC SMART; SM00112; CA; 4.
CC PROSITE; PS00232; CADHERIN 1; 3.
CC PROSITE; PS00268; CADHERIN 2; 4.
CC Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW Signal.
FT SIGNAL 1 25 POTENTIAL.
FT PROPEP 26 99
FT CHAIN 100 822 CADHERIN-3.
FT DOMAIN 100 647 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 648 670 POTENTIAL.
FT DOMAIN 671 822 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 100 207 CADHERIN 1.
FT DOMAIN 208 320 CADHERIN 2.
FT DOMAIN 321 432 CADHERIN 3.
FT DOMAIN 433 538 CADHERIN 4.
FT DOMAIN 539 645 CADHERIN 5.
FT DOMAIN 778 793 SER-RICH.

```

FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 822 AA; 90754 MW; 84975D2115B14DC3 CRC64;

Query Match 6.2%; Score 51; DB 1; Length 822;
 Best Local Similarity 100.0%; Pred. No. 2.5e-39;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 684 KEPLLPEDTRDNVYFGEGBGSDQYDITQLHGLEARPEVVLNDV 734
 DB 677 KEPLLPEDTRDNVYFGEGBGSDQYDITQLHGLEARPEVVLNDV 727

RESULT 4
 ID_CAD3_PIG STANDARD; PRT; 145 AA.
 AC O18926; 2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cadherin-3 (Placental-cadherin) (P-cadherin) (Fragment).
 GN CDH3
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Retinal pigment epithelium;
 RA Lutz D.A., Zheng J.J.
 RT "Expression of multiple cadherins in adult retinal pigment epithelial (RPE) cells."
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins. They preferentially interact with themselves in a homophilic manner in connecting cells; cadherins may thus contribute to the sorting of heterogeneous cell types.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -!- SIMILARITY: Contains 5 cadherin domains.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF033826; AAB87087.1; -.
 HSSP; P09803; 1SUH.
 InterPro; IPR002126; Cadherin.
 Pfam; PF00028; cadherin; 1.
 PRINTS; PR00205; CADHERIN.
 SMART; SM01112; CA; 1.
 PROSITE; PS00232; CADHERIN 1; 1.
 PROSITE; PS0268; CADHERIN 2; 1.
 Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 NON_TER 1
 DOMAIN <1>145 EXTRACELLULAR (POTENTIAL).
 DOMAIN <1>39 CADHERIN 1.
 DOMAIN <1>145 CADHERIN 2.
 CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
 NON_TER 145 145
 SEQUENCE 145 AA; 15617 MW; 8AA2C49E76EB40EC CRC64;

Query Match 3.3%; Score 27; DB 1; Length 145;
 Best Local Similarity 100.0%; Pred. No. 1.8e-17;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 FRGSVLEGLPGLTSMQVATDEDDAI 246
 DB 44 FRGSVLEGLPGLTSMQVATDEDDAI 70

RESULT 5
 ID_CADL_CHICK STANDARD; PRT; 732 AA.
 AC P33145;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE B-cadherin precursor (K-CAM protein) (Fragment).
 GN K-CAM.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92107987; PubMed=1763068;
 RA Sorkin B.C., Gallin W.J., Edelman G.M., Cunningham B.A.;
 RT "Genes for two calcium-dependent cell adhesion molecules have similar structures and are arranged in tandem in the chicken genome."
 RL Proc. Natl. Acad. Sci. U.S.A. 88:11545-11549 (1991).
 RN [2]
 RP SEQUENCE OF 7-732 FROM N.A.
 RC TISSUE=Embryonic brain;
 RX MEDLINE=91225083; PubMed=2026653;
 RA Napolitano E.W., Venstrom K., Wheeler E.F., Reichardt L.F.;
 RT "Molecular cloning and characterization of B-cadherin, a novel chick cadherin."
 RL J. Cell Biol. 113:893-905 (1991).
 CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins. They preferentially interact with themselves in a homophilic manner in connecting cells; cadherins may thus contribute to the sorting of heterogeneous cell types. B-cadherin may have important functions in neurogenesis, in at least some epithelia, and in embryogenesis.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: Expressed in a wide variety of tissues.
 CC -!- SIMILARITY: Contains 5 cadherin domains.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M81894; AAA48929.1; -.
 EMBL; X58518; CAA41408.1; -.
 PIR; A41634; IJCHCB.
 HSSP; P09803; 1SUH.
 InterPro; IPR002126; Cadherin.
 InterPro; IPR00233; Cadherin_C-term.
 Pfam; PF00028; cadherin; 5.
 Pfam; PF01049; Cadherin_C-term; 1.
 PRINTS; PR00205; CADHERIN.
 PROSITE; PS00232; CADHERIN 1; 3.
 PROSITE; PS0268; CADHERIN 2; 5.
 Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
 NON_TER 1
 PROPEP <1>6
 CHAIN 7 732
 DOMAIN 6 554
 TRANSMEM 555 580
 DOMAIN 581 732
 DOMAIN 6 114
 DOMAIN 115 227
 DOMAIN 228 339
 DOMAIN 340 443
 DOMAIN 444 554
 DOMAIN 555 702
 CARBOHYD 137 137
 B-CADHERIN.
 EXTRACELLULAR (POTENTIAL).
 POTENTIAL.
 CYTOPLASMIC (POTENTIAL).
 CADHERIN 1.
 CADHERIN 2.
 CADHERIN 3.
 CADHERIN 4.
 CADHERIN 5.
 SER-RICH.
 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 414 M -> V (IN REF. 2).
SQ SEQUENCE 732 AA; 80613 MW; 091DS9A6A16CBD45 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 1.5e-07; Length 732;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 797 SSASDQDDYDLNHWG 813
Db 699 SSASDQDDYDLNHWG 715

RESULT 6
CAD1_CHICK
ID CAD1_CHICK STANDARD; PRT; 887 AA.
AC P08641;
DT 01-AUG-1998 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epithelial-cadherin precursor (B-cadherin) (Cadherin-1) (Liver cell
DE adhesion molecule) (L-CAM).
GN CDH1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89017248; PubMed=3174655;
RA Sorkin B.C., Hemperly J.J., Edelman G.M., Cunningham B.A.;
RT "Structure of the gene for the liver cell adhesion molecule, L-CAM.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:7617-7621(1988).
RN [2]
RP SEQUENCE OF 51-887 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=87204217; PubMed=3472238;
RA Gallin W.J., Sorkin B.C., Edelman G.M., Cunningham B.A.;
RT "Sequence analysis of a cDNA clone encoding the liver cell adhesion
RT molecule, L-CAM.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:2808-2812(1987).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types. B-cadherin is a ligand for
CC integrin alpha-5/beta-7.
CC -!- SUBUNIT: Homodimer.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Non-neural epithelial tissues.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC -----
DR EMBL; M22190; AAA82572.1; JOINED.
DR EMBL; M22180; AAA82572.1; JOINED.
DR EMBL; M22181; AAA82572.1; JOINED.
DR EMBL; M22182; AAA82572.1; JOINED.
DR EMBL; M22186; AAA82572.1; JOINED.
DR EMBL; M22183; AAA82572.1; JOINED.
DR EMBL; M22194; AAA82572.1; JOINED.
DR EMBL; M22184; AAA82572.1; JOINED.
DR EMBL; M22185; AAA82572.1; JOINED.
DR EMBL; M22189; AAA82572.1; JOINED.
DR EMBL; M22193; AAA82572.1; JOINED.
DR EMBL; M22187; AAA82572.1; JOINED.
DR EMBL; M22192; AAA82572.1; JOINED.
DR EMBL; M22191; AAA82572.1; JOINED.

```

```

DR EMBL; M22195; AAA82572.1; JOINED.
DR EMBL; M1260; AAA82573.1; -.
DR PIR; A30201; IJCHCL.
DR HSSP; P09803; 1SUH.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000333; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin_C_term; 1.
DR PRINTS; P00205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW Signal.
FT SIGNAL 1 26 POTENTIAL.
FT PROPEP 27 160
FT CHAIN 161 887
FT DOMAIN 161 714
FT TRANSMEM 715 735
FT DOMAIN 736 887
FT DOMAIN 161 288
FT DOMAIN 269 381
FT DOMAIN 382 493
FT DOMAIN 494 599
FT DOMAIN 600 704
FT DOMAIN 844 855
FT CARBOHYD 291 291
FT CARBOHYD 346 346
FT CARBOHYD 564 564
FT CARBOHYD 643 643
FT CONFLICT 140 140
SQ SEQUENCE 887 AA; 97783 MW; F56A8A9779A94A40 CRC64;

```

Query Match 2.1%; Score 17; DB 1; Length 887;
 Best Local Similarity 100.0%; Pred. No. 1.8e-07;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 797 SSASDQDDYDLNHWG 813
Db 854 SSASDQDDYDLNHWG 870

```

RESULT 7

```

ID CADF_XENLA STANDARD; PRT; 880 AA.
AC P33148; O91543;
DT 01-OCT-1993 (Rel. 27, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE EP-cadherin precursor (C-cadherin).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91372132; PubMed=1893866;
RA Ginsberg D., Deslmon D., Geiger B.;
RT "Expression of a novel cadherin (EP-cadherin) in unfertilized eggs
RT and early Xenopus embryos.";
RN [2]
RP SEQUENCE FROM N.A.
RA Lee C.H.;
RN [3]
RP REVISION TO 260.
RA Flament S.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the

```

CC sorting of heterogeneous cell types.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DEVELOPMENTAL STAGE: Unfertilized eggs and early xenopus embryos.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X63720; CAA45252.1; ALT_INIT.
CC EMBL; U04707; AAC16910.1; -.
CC DR; PIR; B43785; IUXLCP.
CC DR; PIR; 1L3W; 26-APR-02.
CC DR; PIR; 1L3W; 26-APR-02.
CC DR; InterPro; IPR002126; Cadherin.
CC DR; InterPro; IPR000233; Cadherin_C_term.
CC DR; Pfam; PF00028; cadherin; 5.
CC DR; Pfam; PF01049; Cadherin_C_term; 1.
CC DR; PRINTS; PR00205; CADHERIN.
CC DR; SMART; SM00112; CA; 4.
CC DR; PROSITE; PS00232; CADHERIN 1; 3.
CC DR; PROSITE; PS0268; CADHERIN 2; 4.
CC DR; Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
CC KW Signal; 3D-structure.
CC FT SIGNAL 1 28 POTENTIAL.
CC FT PROPEP 29 155 POTENTIAL.
CC FT CHAIN 156 880 EP-CADHERIN.
CC FT DOMAIN 156 703 EXTRACELLULAR (POTENTIAL).
CC FT TRANSSEM 704 728 POTENTIAL.
CC FT DOMAIN 729 880 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 156 263 CADHERIN 1.
CC FT DOMAIN 264 376 CADHERIN 2.
CC FT DOMAIN 377 487 CADHERIN 3.
CC FT DOMAIN 488 593 CADHERIN 4.
CC FT DOMAIN 594 704 CADHERIN 5.
CC FT DOMAIN 837 850 SER-RICH.
CC FT DOMAIN 876 880 ASP/GLU-RICH (ACIDIC).
CC FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 681 681 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CONFLICT 3 3 S -> G (IN REF. 2).
CC FT CONFLICT 112 112 I -> N (IN REF. 2).
CC FT CONFLICT 573 573 S -> P (IN REF. 2).
CC FT CONFLICT 864 864 S -> P (IN REF. 2).
CC FT CONFLICT 864 864 S -> P (IN REF. 2).
CC SQ SEQUENCE 880 AA; 97650 MW; C6CCD91566427D86 CRC64;
Query Match 1.9%; Score 16; DB 1; Length 880;
Best Local Similarity 100.0%; Pred. No. 1.5e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 699 FYGEGGGEDQDYD 714
Db 750 FYGEGGGEDQDYD 765
RESULT 8
ID CADB_XENLA STANDARD; PRT; 884 AA.
AC P33152;
DT 01-OCT-1993 (Rel. 27, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Blastomere-cadherin precursor (B-cadherin) (XBCad).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=95151580; PubMed=7531482;
RA Mueller H.-A.J.; Kuehl M.; Finnemann S.; Schneider S.,
van der Poel S.Z.; Hausen P.; Wedlich D.;
RT "Xenopus cadherins: the maternal pool comprises distinguishable
members of the family";
RL Mech. Dev. 47:213-223(1994).
RN [2]
RP SEQUENCE OF 438-884 FROM N.A.
RX MEDLINE=92062581; PubMed=1840622;
RA Herzberg P.; Wildermuth V.; Wedlich D.;
RT "Expression of XBCad, a novel cadherin, during oogenesis and early
development of Xenopus";
RL Mech. Dev. 35:33-42(1991).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Expressed in pituitary gland, lung and kidney.
CC -!- DEVELOPMENTAL STAGE: During oogenesis and early development.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X78546; CAA55252.1; ALT_INIT.
CC EMBL; X63719; CAA45251.1; -.
CC HSP; P09803; ISUH.
CC DR; InterPro; IPR002126; Cadherin.
CC DR; InterPro; IPR000233; Cadherin_C_term.
CC DR; Pfam; PF00028; cadherin; 5.
CC DR; Pfam; PF01049; Cadherin_C_term; 1.
CC DR; PRINTS; PR00205; CADHERIN.
CC DR; SMART; SM00112; CA; 4.
CC DR; PROSITE; PS00232; CADHERIN 1; 3.
CC DR; PROSITE; PS0268; CADHERIN 2; 4.
CC DR; Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
CC KW Signal.
CC FT SIGNAL 1 26 POTENTIAL.
CC FT PROPEP 27 157 POTENTIAL.
CC FT CHAIN 158 884 BLASTOMERE-CADHERIN.
CC FT DOMAIN 158 706 EXTRACELLULAR (POTENTIAL).
CC FT TRANSSEM 707 730 POTENTIAL.
CC FT DOMAIN 731 884 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 158 265 CADHERIN 1.
CC FT DOMAIN 266 378 CADHERIN 2.
CC FT DOMAIN 379 489 CADHERIN 3.
CC FT DOMAIN 490 595 CADHERIN 4.
CC FT DOMAIN 596 706 CADHERIN 5.
CC FT DOMAIN 828 852 SER-RICH.
CC FT DOMAIN 879 884 ASP/GLU-RICH (ACIDIC).
CC FT CARBOHYD 427 427 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 560 560 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 683 683 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CONFLICT 438 440 ILT -> NSA (IN REF. 2).
CC FT CONFLICT 677 677 R -> Q (IN REF. 2).
CC FT CONFLICT 787 787 V -> A (IN REF. 2).
CC FT CONFLICT 820 820 D -> N (IN REF. 2).
CC FT CONFLICT 857 857 D -> N (IN REF. 2).
CC FT CONFLICT 863 863 D -> N (IN REF. 2).
CC FT CONFLICT 882 882 MISSING (IN REF. 2).
CC SQ SEQUENCE 884 AA; 97980 MW; 9B865D7E1DCB75B CRC64;
Query Match 1.9%; Score 16; DB 1; Length 884;
Best Local Similarity 100.0%; Pred. No. 1.5e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

QY 699 FYYGEGGGEDDDYD 714
Db 752 FYYGEGGGEDDDYD 767

RESULT 9
CAD1_XENLA STANDARD; PRT; 872 AA.
AC P30944; Q91709;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epithelial-cadherin precursor (E-cadherin) (Uvomorulin) (XTCAD-1).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
CX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95324376; PubMed=7600966;
RA Levine E., Lee C.H., Kintner C., Gumbiner B.M.;
RT "Selective disruption of E-cadherin function in early Xenopus embryos
RL by a dominant negative mutant.";
RL Development 120:901-909(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=tail bud;
RX MEDLINE=94368839; PubMed=8086449;
RA Tooi O., Fujii G., Tashiro K., Shikawa K.;
RT "Molecular cloning of cDNA for XTCAD-1, a novel Xenopus cadherin, and
RL its expression in adult tissues and embryos of Xenopus laevis.";
RL Biochim. Biophys. Acta 1219:121-128(1994).
RN [3]
RP SEQUENCE OF 149-872 FROM N.A.
RX MEDLINE=94363396; PubMed=8081882;
RA Broders F., Girault J.M., Simonneau L., Thiery J.P.;
RT "Sequence and distribution of Xenopus laevis E-cadherin transcripts.";
RL Cell Adhes. Commun. 1:265-277(1993).
RN [4]
RP SEQUENCE OF 149-872.
RX MEDLINE=91347911; PubMed=1879345;
RA Angres E., Mueller A.H.J., Kellermann J., Hausen P.;
RT "Differential expression of two cadherins in Xenopus laevis.";
RL Development 111:829-844(1991).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types. E-cadherin is a ligand for
CC integrin alpha-E/beta-7.
CC -!- SUBUNIT: Homodimer.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Non-neural epithelial tissues.
CC -!- DEVELOPMENTAL STAGE: Appears in the embryonic ectoderm during
CC gastrulation when epidermal differentiation commences and it
CC disappears from the neural plate area upon neural induction.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U04708; AAA93116.1; -
CC EMBL; L29057; AAA61489.1; -
CC EMBL; X75454; CAA53206.1; -
CC HSSP; P09803; ISUH.
CC InterPro; IPR002126; Cadherin.
CC InterPro; IPR002233; Cadherin_C_term.
CC Pfam; PF00028; cadherin; 5.

```

```

DR Pfam; PF01049; Cadherin_C_term; 1.
DR PRINTS; PRO0205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS50268; CADHERIN_2; 4.
KW Cell adhesion; Glycoprotein; transmembrane; Calcium-binding; Repeat;
KW Signal.
FT SIGNAL. 1 25
FT PROPEP 26 148
FT CHAIN 149 872
FT DOMAIN 149 701
FT TRANSMEM 702 722
FT DOMAIN 723 872
FT DOMAIN 244 254
FT DOMAIN 358 368
FT DOMAIN 577 587
FT DOMAIN 710 721
FT CARBOHYD 209 209
FT CARBOHYD 456 456
FT CARBOHYD 552 552
FT CARBOHYD 631 631
FT CARBOHYD 669 669
FT CONFLICT 242 242
FT CONFLICT 332 332
FT CONFLICT 487 487
FT CONFLICT 501 502
FT CONFLICT 539 541
FT CONFLICT 548 548
FT CONFLICT 557 557
FT CONFLICT 567 567
FT CONFLICT 602 604
FT CONFLICT 638 640
FT CONFLICT 640 647
FT CONFLICT 660 660
FT CONFLICT 770 770
FT CONFLICT 842 842
FT CONFLICT 870 871
SQ SEQUENCE 872 AA; 96065 MW; 08637967EEFB4664 CRC64;

```

Query Match 1.8%; Score 15; DB 1; Length 872;

Best Local Similarity 100.0%; Pred. No. 1.3e-05; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 744 YRPRPNPDEIGNFI 758

Db 788 YRPRPNPDEIGNFI 802

RESULT 10

```

CAD1_HUMAN
ID CAD1_HUMAN STANDARD; PRT; 882 AA.
AC P12830; Q13799; Q14216; Q15855; Q16194;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-JUL-1993 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Epithelial-cadherin precursor (E-cadherin) (Uvomorulin) (Cadherin-1)
DE (CAM 120/80).
GN CDH1 OR UVO OR CDHE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Pancreas;
RX MEDLINE=93211394; PubMed=8459805;
RA Bussemakers M.J.G., Mees S.G.M., van Bokhoven A., Debruyne F.M.J.,
RA Schaiken J.A.;
RT "Molecular cloning and characterization of the human E-cadherin
RL cDNA.";
RL Mol. Biol. Rep. 17:123-128(1993).
RN [2]
RP SEQUENCE FROM N.A.

```

RA Kelker W., Warda A., Oda T., Hirohashi S., Kemler R., Birchmeier W.;
 RA Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Rimm D.L., Morrow J.S.;
 RA "Molecular cloning of human E-cadherin suggests a novel subdivision of
 the cadherin superfamily";
 RT Biochem. Biophys. Res. Commun. 200:1754-1761(1994).
 RN [4]
 RN SEQUENCE OF 56-882 FROM N.A.
 RC TISSUE=Placenta;
 RA MEDLINE=9532420; PubMed=8185635;
 RA Rimm D.L., Morrow J.S.;
 RA "Molecular cloning of human E-cadherin suggests a novel subdivision of
 the cadherin superfamily";
 RT Biochem. Biophys. Res. Commun. 200:1754-1761(1994).
 RN [4]
 RN SEQUENCE OF 172-311 FROM N.A.
 RC TISSUE=Liver;
 RA MEDLINE=89031725; PubMed=3263290;
 RA Mansouri A., Spurr N., Goodfellow P.N., Kemler R.;
 RA "Characterization and chromosomal localization of the gene encoding
 the human cell adhesion molecule uvomorulin";
 RT Differentiation 38:67-71(1988).
 RN [6]
 RN SEQUENCE OF 265-392 FROM N.A.
 RC TISSUE=Liver;
 RA Frixen U.H.;
 RA Submitted (MAR-1990) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RN SEQUENCE OF 1-16 FROM N.A.
 RC MEDLINE=94380041; PubMed=8093045;
 RA Bussemakers M.J.G., Girolodi L.A., van Bokhoven A., Schalken J.A.;
 RA "Transcriptional regulation of the human E-cadherin gene in human
 prostate cancer cell lines: characterization of the human E-cadherin
 gene promoter";
 RT Biochem. Biophys. Res. Commun. 203:1284-1290(1994).
 RN [8]
 RN SEQUENCE OF 1-16 FROM N.A.
 RC TISSUE=Placenta;
 RA MEDLINE=95365379; PubMed=7543680;
 RA Yoshiura K., Kanai Y., Ochiai A., Shimoyama Y., Sugimura T.,
 RA Hirohashi S.;
 RA "Silencing of the E-cadherin invasion-suppressor gene by CpG
 methylation in human carcinomas";
 RT Proc. Natl. Acad. Sci. U.S.A. 92:7416-7419(1995).
 RN [9]
 RN INTERCHAIN DISULFIDE BOND.
 RA MEDLINE=21975237; PubMed=11856785;
 RA Makgansar I.T., Nguyen P.D., Ikseue A., Kuczerka K., Dentler W.,
 RA Urbauer J.L., Galeva N., Alterman M., Sahaan T.J.;
 RA "Disulfide bond formation promotes the cis- and trans-dimerization of
 the E-cadherin-derived first repeat";
 RT J. Biol. Chem. 277:16002-16010(2002).
 RN [10]
 RN DISEASE
 RA MEDLINE=99406500; PubMed=10477433;
 RA Guilford P.J., Hopkins J.B.W., Grady W.M., Markowitz S.D., Willis J.,
 RA Lynch H., Rajput A., Wiesner G.L., Lindor N.M., Burgart L.J.,
 RA Toro T.T., Lee D., Limacher J.-M., Shaw D.W., Findlay M.P.N.,
 RA Reeve A.E.;
 RA "E-cadherin germline mutations define an inherited cancer syndrome
 dominated by diffuse gastric cancer";
 RT Hum. Mutat. 14:249-255(1999).
 RN [11]
 RN REVIEW ON VARIANTS.
 RA MEDLINE=98415721; PubMed=9744472;
 RA Berr G., Becker K.-F., Hoefler H., van Roy F.;
 RA "Mutations of the human E-cadherin (CDH1) gene";
 RT Hum. Mutat. 12:226-237(1998).
 RN [12]
 RN SEQUENCE OF 337-476 FROM N.A., AND VARIANTS ALA-370 AND ASP-473.
 RA MEDLINE=94306394; PubMed=8033105;
 RA Becker K.-F., Atkinson M.J., Reich U., Becker I., Nekarda H.,
 RA Sievert J.R., Hoefler H.;
 RA "E-cadherin gene mutations provide clues to diffuse type gastric
 carcinomas";
 RT Cancer Res. 54:3845-3852(1994).
 RN [13]
 RN VARIANT LOBULAR BREAST CARCINOMA SER-315.
 RA MEDLINE=95049851; PubMed=7961105;
 RA Kanai Y., Oda T., Tsuda H., Ochiai A., Hirohashi S.;
 RA "Point mutation of the E-cadherin gene in invasive lobular carcinoma
 of the breast";
 RT Jpn. J. Cancer Res. 85:1035-1039(1994).
 RN [14]
 RN VARIANTS GYNECOLOGIC CANCERS THR-617; VAL-711 AND GLY-838.
 RA MEDLINE=94355985; PubMed=8075649;
 RA Risinger J.I., Berchuck A., Kohler M.F., Boyd J.;
 RA "Mutations of the E-cadherin gene in human gynecologic cancers";
 RT Nat. Genet. 7:98-102(1994).
 RN [15]
 RN VARIANT GASTRIC ADENOCARCINOMA 274-GLY--PRO-277 DEL.
 RA MEDLINE=94173928; PubMed=8127895;
 RA Oda T., Kanai Y., Oyama T., Yoshiura K., Shimoyama Y., Birchmeier W.,
 RA Sugimura T., Hirohashi S.;
 RA "E-cadherin gene mutations in human gastric carcinoma cell lines";
 RT Proc. Natl. Acad. Sci. U.S.A. 91:1858-1862(1994).
 RN [16]
 RN VARIANT GASTRIC CARCINOMA PRO-193.
 RA MEDLINE=96390918; PubMed=8797891;
 RA Muta H., Noguchi M., Kanai Y., Ochiai A., Nawata H., Hirohashi S.;
 RA "E-cadherin gene mutations in signet ring cell carcinoma of the
 stomach";
 RT Jpn. J. Cancer Res. 87:843-848(1996).
 RN [17]
 RN VARIANTS GASTRIC CARCINOMA ASP-400 DEL AND 418-ASP--PHE-423 DEL.
 RA MEDLINE=97197648; PubMed=9045944;
 RA Tamura G., Sakata K., Nishizuka S., Maesawa C., Suzuki Y., Iwaya T.,
 RA Terashima M., Saito K., Satodate R.;
 RA "Inactivation of the E-cadherin gene in primary gastric carcinomas and
 gastric carcinoma cell lines";
 RT Jpn. J. Cancer Res. 87:1153-1159(1996).
 RN [18]
 RN VARIANT THYROID TUMOR THR-592.
 RA MEDLINE=97138061; PubMed=8985087;
 RA Soares P., Berr G., van Roy F., Sobrinho-Simoes M.;
 RA "E-cadherin gene alterations are rare events in thyroid tumors";
 RT Int. J. Cancer 70:32-38(1997).
 RN [19]
 RN VARIANTS ASP-336 AND ILE-470.
 RA MEDLINE=98196671; PubMed=9537325;
 RA Guilford P.J., Hopkins J.B.W., Harraway J., McLeod M., McLeod N.,
 RA Harawira P., Taite H., Scoular R., Miller A., Reeve A.E.;
 RA "E-cadherin germline mutations in familial gastric cancer";
 RT Nature 392:402-405(1998).
 RN [20]
 RN VARIANTS HDGC GLY-244 AND ALA-487.
 RA MEDLINE=99253140; PubMed=10319582;
 RA Yoon K.-A., Xu J.-L., Yang H.-K., Kim W.H., Park S.Y., Park J.-G.;
 RA "Germline mutations of E-cadherin gene in Korean familial gastric
 cancer patients";
 RT J. Hum. Genet. 44:177-180(1999).
 RN [21]
 RN VARIANT COLORECTAL CANCER ALA-340.
 RA MEDLINE=20357134; PubMed=10895919;
 RA Kim H.C., Wheeler J.M.D., Kim J.C., Ilyas M., Beck N.E., Kim B.S.,
 RA Park K.C., Bodmer W.F.;
 RA "The E-cadherin gene (CDH1) variants T340A and L599V in gastric and
 colorectal cancer patients in Korea";
 RT Gut 47:262-267(2000).
 RN [22]
 RN VARIANT ALA-270.


```

FT STRAND 282 283 284 285 286 287 288 289
FT STRAND 288 289 290 291 292 293 294 295
FT TURN 294 295 296 297 298 299 300 301
FT TURN 299 300 301 302 303 304 305 306
FT STRAND 303 304 305 306 307 308 309 310
FT STRAND 307 308 309 310 311 312 313 314
FT TURN 310 311 312 313 314 315 316 317
FT TURN 312 313 314 315 316 317 318 319
FT STRAND 314 315 316 317 318 319 320 321
FT STRAND 317 318 319 320 321 322 323 324
FT TURN 320 321 322 323 324 325 326 327
FT TURN 322 323 324 325 326 327 328 329
FT STRAND 324 325 326 327 328 329 330 331
FT STRAND 327 328 329 330 331 332 333 334
FT TURN 330 331 332 333 334 335 336 337
FT TURN 332 333 334 335 336 337 338 339
FT STRAND 334 335 336 337 338 339 340 341
FT STRAND 337 338 339 340 341 342 343 344
FT TURN 340 341 342 343 344 345 346 347
FT TURN 342 343 344 345 346 347 348 349
FT STRAND 344 345 346 347 348 349 350 351
FT STRAND 347 348 349 350 351 352 353 354
FT TURN 350 351 352 353 354 355 356 357
FT TURN 352 353 354 355 356 357 358 359
FT STRAND 354 355 356 357 358 359 360 361
FT STRAND 357 358 359 360 361 362 363 364
FT STRAND 360 361 362 363 364 365 366 367
FT STRAND 363 364 365 366 367 368 369 370
FT STRAND 366 367 368 369 370 371 372 373
FT STRAND 369 370 371 372 373 374 375 376
FT STRAND 372 373 374 375 376 377 378 379
FT STRAND 375 376 377 378 379 380 381 382
FT STRAND 378 379 380 381 382 383 384 385
FT STRAND 381 382 383 384 385 386 387 388
FT STRAND 384 385 386 387 388 389 390 391
FT STRAND 387 388 389 390 391 392 393 394
FT STRAND 390 391 392 393 394 395 396 397
FT STRAND 393 394 395 396 397 398 399 400
FT STRAND 396 397 398 399 400 401 402 403
FT STRAND 399 400 401 402 403 404 405 406
FT STRAND 402 403 404 405 406 407 408 409
FT STRAND 405 406 407 408 409 410 411 412
FT STRAND 408 409 410 411 412 413 414 415
FT STRAND 411 412 413 414 415 416 417 418
FT STRAND 414 415 416 417 418 419 420 421
FT STRAND 417 418 419 420 421 422 423 424
FT STRAND 420 421 422 423 424 425 426 427
FT STRAND 423 424 425 426 427 428 429 430
FT STRAND 426 427 428 429 430 431 432 433
FT STRAND 429 430 431 432 433 434 435 436
FT STRAND 432 433 434 435 436 437 438 439
FT STRAND 435 436 437 438 439 440 441 442
FT STRAND 438 439 440 441 442 443 444 445
FT STRAND 441 442 443 444 445 446 447 448
FT STRAND 444 445 446 447 448 449 450 451
FT STRAND 447 448 449 450 451 452 453 454
FT STRAND 450 451 452 453 454 455 456 457
FT STRAND 453 454 455 456 457 458 459 460
FT STRAND 456 457 458 459 460 461 462 463
FT STRAND 459 460 461 462 463 464 465 466
FT STRAND 462 463 464 465 466 467 468 469
FT STRAND 465 466 467 468 469 470 471 472
FT STRAND 468 469 470 471 472 473 474 475
FT STRAND 471 472 473 474 475 476 477 478
FT STRAND 474 475 476 477 478 479 480 481
FT STRAND 477 478 479 480 481 482 483 484
FT STRAND 480 481 482 483 484 485 486 487
FT STRAND 483 484 485 486 487 488 489 490
FT STRAND 486 487 488 489 490 491 492 493
FT STRAND 489 490 491 492 493 494 495 496
FT STRAND 492 493 494 495 496 497 498 499
FT STRAND 495 496 497 498 499 500 501 502
FT STRAND 498 499 500 501 502 503 504 505
FT STRAND 501 502 503 504 505 506 507 508
FT STRAND 504 505 506 507 508 509 510 511
FT STRAND 507 508 509 510 511 512 513 514
FT STRAND 510 511 512 513 514 515 516 517
FT STRAND 513 514 515 516 517 518 519 520
FT STRAND 516 517 518 519 520 521 522 523
FT STRAND 519 520 521 522 523 524 525 526
FT STRAND 522 523 524 525 526 527 528 529
FT STRAND 525 526 527 528 529 530 531 532
FT STRAND 528 529 530 531 532 533 534 535
FT STRAND 531 532 533 534 535 536 537 538
FT STRAND 534 535 536 537 538 539 540 541
FT STRAND 537 538 539 540 541 542 543 544
FT STRAND 540 541 542 543 544 545 546 547
FT STRAND 543 544 545 546 547 548 549 550
FT STRAND 546 547 548 549 550 551 552 553
FT STRAND 549 550 551 552 553 554 555 556
FT STRAND 552 553 554 555 556 557 558 559
FT STRAND 555 556 557 558 559 560 561 562
FT STRAND 558 559 560 561 562 563 564 565
FT STRAND 561 562 563 564 565 566 567 568
FT STRAND 564 565 566 567 568 569 570 571
FT STRAND 567 568 569 570 571 572 573 574
FT STRAND 570 571 572 573 574 575 576 577
FT STRAND 573 574 575 576 577 578 579 580
FT STRAND 576 577 578 579 580 581 582 583
FT STRAND 579 580 581 582 583 584 585 586
FT STRAND 582 583 584 585 586 587 588 589
FT STRAND 585 586 587 588 589 590 591 592
FT STRAND 588 589 590 591 592 593 594 595
FT STRAND 591 592 593 594 595 596 597 598
FT STRAND 594 595 596 597 598 599 600 601
FT STRAND 597 598 599 600 601 602 603 604
FT STRAND 600 601 602 603 604 605 606 607
FT STRAND 603 604 605 606 607 608 609 610
FT STRAND 606 607 608 609 610 611 612 613
FT STRAND 609 610 611 612 613 614 615 616
FT STRAND 612 613 614 615 616 617 618 619
FT STRAND 615 616 617 618 619 620 621 622
FT STRAND 618 619 620 621 622 623 624 625
FT STRAND 621 622 623 624 625 626 627 628
FT
```

```
CC EMBL; AB017696; BAA84920.1; -
CC EMBL; AF177680; AAR87055.1; -
CC EMBL; AJ000540; CAA04173.1; -
CC HSPF; P03903; IEDH.
CC InterPro; IPR002126; Cadherin.
CC InterPro; IPR00233; Cadherin_C_term.
CC Pfam; PF00028; cadherin; 5.
CC Pfam; PF01049; Cadherin_C_term; 1.
CC PRINTS; PR00205; CADHERIN.
CC SMART; SM00112; CA; 4.
CC PROSITE; PS00232; CADHERIN_1; 3.
CC PROSITE; PS0268; CADHERIN_2; 5.
CC Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW Signal; Phosphorylation.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 158 POTENTIAL.
FT CHAIN 159 886 EPIHELIAL-CADHERIN.
FT DOMAIN 24 713 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 714 734 POTENTIAL.
FT DOMAIN 735 886 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1 266 CADHERIN_1.
FT DOMAIN 267 379 CADHERIN_2.
FT DOMAIN 380 490 CADHERIN_3.
FT DOMAIN 491 597 CADHERIN_4.
FT DOMAIN 598 701 CADHERIN_5.
FT DOMAIN 842 857 SER-RICH.
FT MOD_RES 842 842 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 844 844 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 850 850 PHOSPHORYLATION (BY SIMILARITY).
FT CARBOHYD 562 562 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 641 641 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 886 AA; 98714 MW; A9AE2BE8797A547 CRC64;

Query Match 1.8%; Score 15; DB 1; Length 886;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 744 YRPRNPANPEIGNFI 758
Db 801 YRPRNPANPEIGNFI 815

RESULT 13
CAD4_CHICK
ID CAD4_CHICK STANDARD; PRT; 913 AA.
AC F24503;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cadherin-4 precursor (Retinal-cadherin) (R-cadherin) (R-CAD).
GN CDH4.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
RN [1]_TaxID=9031;
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=91299341; PubMed=1712604;
RA Inuzuka H., Miyatani S., Takeichi M.;
RT "R-cadherin: a novel Ca(2+)-dependent cell-cell adhesion molecule
expressed in the retina.";
RL Neuron 7:69-79(1991).
CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
CC They preferentially interact with themselves in a homophilic
CC manner in connecting cells; cadherins may thus contribute to the
CC sorting of heterogeneous cell types. May play an important role in
CC retinal development.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Embryonic brain and neuronal retina.
CC -!- DEVELOPMENTAL STAGE: Detected only after some degree of neuronal
```

```
CC differentiation has taken place and persists at least up to the
CC newly hatched stage.
CC -!- SIMILARITY: Contains 5 cadherin domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D14459; BAA03356.1; -
CC PIR; JH0424; ICHCR.
CC HSPF; P15116; INCI.
CC InterPro; IPR002126; Cadherin.
CC InterPro; IPR00233; Cadherin_C_term.
CC Pfam; PF00028; cadherin; 5.
CC Pfam; PF01049; Cadherin_C_term; 1.
CC PRINTS; PR00205; CADHERIN.
CC SMART; SM00112; CA; 5.
CC PROSITE; PS00232; CADHERIN_1; 3.
CC PROSITE; PS0268; CADHERIN_2; 5.
CC Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW Signal.
FT SIGNAL 1 19 POTENTIAL.
FT PROPEP 20 166 POTENTIAL.
FT CHAIN 167 913 CADHERIN-4.
FT DOMAIN 167 731 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 732 753 POTENTIAL.
FT DOMAIN 754 913 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 167 274 CADHERIN_1.
FT DOMAIN 275 389 CADHERIN_2.
FT DOMAIN 390 504 CADHERIN_3.
FT DOMAIN 505 610 CADHERIN_4.
FT DOMAIN 611 721 CADHERIN_5.
FT DOMAIN 870 885 SER-RICH.
FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 554 554 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 658 658 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 699 699 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 652 652 P -> T (IN ONE FORM).
SQ SEQUENCE 913 AA; 100885 MW; BD2BB9ECC815DD6 CRC64;

Query Match 1.6%; Score 13; DB 1; Length 913;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 RFKKLADMYGGGE 827
Db 899 RFKKLADMYGGGE 911

RESULT 14
CAD4_MOUSE
ID CAD4_MOUSE STANDARD; PRT; 913 AA.
AC P39038;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Cadherin-4 precursor (Retinal-cadherin) (R-cadherin) (R-CAD).
GN CDH4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreatic islets;
RX MEDLINE=94067164; PubMed=8247017;
RA Hutton J.C., Christofori G., Chi W.Y., Edman U., Guest P.C.,
```

RA Hanahan D., Kelly R.B.;
 RT "Molecular cloning of mouse pancreatic islet R-cadherin: differential
 RT expression in endocrine and exocrine tissue.";
 RL Mol. Endocrinol. 7:1151-1160(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94095672; PubMed=8270638;
 RA Matsunami H., Miyatani S., Inoue T., Copeland N.G., Gilbert D.,
 RA Jenkins N.A., Takeichi M.;
 RT "Cell binding specificity of mouse R-cadherin and chromosomal mapping
 RT of the gene.";
 RL J. Cell Sci. 106:401-409(1993).
 CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
 CC They preferentially interact with themselves in a homophilic
 CC manner in connecting cells; cadherins may thus contribute to the
 CC sorting of heterogeneous cell types. May play an important role
 CC in retinal development.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: Distributed widely in mouse tissues with high
 CC levels present in brain, skeletal muscle and thymus.
 CC -!- SIMILARITY: Contains 5 cadherin domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X69966; CAA49589.1; -;
 DR EMBL; D14888; BAA03605.1; -;
 DR PIR; A47543; A47543;
 DR HSSP; P15116; INCI.
 DR MGD; MGI:99218; Cdh4.
 DR InterPro; IPR002126; Cadherin.
 DR InterPro; IPR002233; Cadherin_C_term.
 DR Pfam; PF00028; cadherin; 5.
 DR Pfam; PF01049; Cadherin_C_term; 1.
 DR PRINTS; PRO0205; CADHERIN.
 DR SMART; SM00112; CA; 5.
 DR PROSITE; PS00232; CADHERIN_1; 3.
 DR PROSITE; PS0268; CADHERIN_2; 5.
 DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
 KW Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT PROPEP 21 166 POTENTIAL.
 FT CHAIN 167 913 CADHERIN-4.
 FT DOMAIN 167 731 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 732 753 POTENTIAL.
 FT DOMAIN 754 913 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 167 274 CADHERIN_1.
 FT DOMAIN 275 389 CADHERIN_2.
 FT DOMAIN 390 504 CADHERIN_3.
 FT DOMAIN 505 610 CADHERIN_4.
 FT DOMAIN 611 721 CADHERIN_5.
 FT DOMAIN 870 885 SER-RICH.
 FT CARBOHYD 146 146 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 280 280 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 409 409 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 554 554 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 658 658 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 699 699 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 913 AA; 10030 MW; 1245A8CE8C338481 CRC64;
 Query Match 1.6%; Score 13; DB 1; Length 913;
 Best Local Similarity 100.0%; Pred. No. 0.001;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 815 RFKGLADMYGGGE 827
 DB 899 RFKGLADMYGGGE 911

RESULT 15.
 CAD4 HUMAN
 ID CAD4 HUMAN STANDARD; PRT; 916 AA.
 AC P55283;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cadherin-4 precursor (Retinal-cadherin) (R-cadherin) (R-CAD).
 DE CDH4.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain.
 RX MEDLINE=95073006; PubMed=7982033;
 RA Tanihara H., Sano K., Helmark R.L., St John T., Suzuki S.;
 RT "Cloning of five human cadherins clarifies characteristic features of
 RT cadherin extracellular domain and provides further evidence for two
 RT structurally different types of cadherin.";
 RL Cell Adhes. Commun. 2:15-26(1994).
 RN [2]
 RP SEQUENCE OF 393-916 FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=91283540; PubMed=2059658;
 RA Suzuki S., Sano K., Tanihara H.;
 RT "Diversity of the cadherin family: evidence for eight new cadherins
 RT in nervous tissue.";
 RL Cell Regul. 2:261-270(1991).
 CC -!- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
 CC They preferentially interact with themselves in a homophilic
 CC manner in connecting cells; cadherins may thus contribute to the
 CC sorting of heterogeneous cell types. May play an important role
 CC in retinal development.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: Expressed mainly in brain but also found in
 CC other tissues.
 CC -!- SIMILARITY: Contains 5 cadherin domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; L34059; AAA35627.1; -;
 DR PIR; C38992; C38992.
 DR HSSP; P15116; INCI.
 DR Genew; HGNC:1763; CDH4.
 DR MIM; 603006; -;
 DR GO; GO:0007155; P:cell adhesion; TAS.
 DR InterPro; IPR002126; Cadherin.
 DR InterPro; IPR002233; Cadherin_C_term.
 DR Pfam; PF00028; cadherin; 5.
 DR Pfam; PF01049; Cadherin_C_term; 1.
 DR SMART; SM00112; CA; 5.
 DR PROSITE; PS00232; CADHERIN_1; 2.
 DR PROSITE; PS0268; CADHERIN_2; 5.
 DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
 KW Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT PROPEP 21 169 POTENTIAL.
 FT CHAIN 170 916 CADHERIN-4.
 FT DOMAIN 170 734 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 735 756 POTENTIAL.
 FT DOMAIN 757 916 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 170 277 CADHERIN_1.

FT DOMAIN 278 392 CADHERIN 2.
FT DOMAIN 393 507 CADHERIN 3.
FT DOMAIN 508 613 CADHERIN 4.
FT DOMAIN 614 724 CADHERIN 5.
FT DOMAIN 873 888 SER-RICH.
FT CARBOHYD 283 283 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 412 412 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 632 632 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 661 661 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 916 AA; 100446 MW; E717C54A19E0C52A CRC64;

Query Match 1.6%; Score 13; DB 1; Length 916;

Best Local Similarity 100.0%; Pred. No. 0.001;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 RFKKLADMYGGGE 827

Db 902 RFKKLADMYGGGE 914

Search completed: September 21, 2004, 22:25:56
Job time : 27 secs

This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:10:15 ; Search time 136 Seconds
(without alignments)
1923.269 Million cell updates/sec

Title: US-09-916-849A-1
Perfect score: 829
Sequence: 1 MGLPRGLASLLLLQVCWLQ.....NEWGRFKKLADMYGGEDD 829

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL 25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51	6.2	753	11 Q80VY6	Q80VY6 mus musculus
2	51	6.2	821	11 Q8BRE1	Q8BRE1 mus musculus
3	51	6.2	822	11 Q8BSL6	Q8BSL6 mus musculus
4	37	4.5	105	11 Q9JIV6	Q9JIV6 rattus norv
5	15	1.8	821	4 Q9JII8	Q9JII8 homo sapien
6	15	1.8	901	4 Q9JII7	Q9JII7 homo sapien
7	14	1.7	40	6 Q18222	Q18222 sus scrofa
8	14	1.7	864	13 Q90Z37	Q90Z37 brachydanio
9	14	1.7	2909	5 Q8MY78	Q8MY78 asterina pe
10	13	1.6	566	4 Q9HIG6	Q9HIG6 homo sapien
11	13	1.6	798	5 Q9NL47	Q9NL47 ciona intes
12	13	1.6	824	4 Q8NB64	Q8NB64 homo sapien
13	12	1.4	68	11 Q80XX1	Q80XX1 mus sp. n-c
14	12	1.4	140	4 Q8ND09	Q8ND09 homo sapien
15	12	1.4	490	13 Q9W6G5	Q9W6G5 brachydanio
16	12	1.4	906	4 Q8N173	Q8N173 homo sapien

17	12	1.4	906	11 Q8BSI9	Q8BSI9 mus musculu
18	12	1.4	922	13 P79883	P79883 xenopus. ma
19	11	1.3	111	11 Q9JIV7	Q9JIV7 rattus norv
20	11	1.3	792	13 Q9DFS1	Q9DFS1 xenopus lae
21	11	1.3	792	13 Q9DFS0	Q9DFS0 xenopus lae
22	10	1.2	66	6 Q8HX20	Q8HX20 bos taurus
23	10	1.2	177	12 Q8QVM3	Q8QVM3 hepatitis e
24	10	1.2	674	12 Q7TGF1	Q7TGF1 hepatitis e
25	10	1.2	1959	5 Q8MY75	Q8MY75 ptychodera
26	9	1.1	105	11 Q9JII1	Q9JII1 rattus norv
27	9	1.1	124	16 Q82ET9	Q82ET9 streptomyce
28	9	1.1	162	11 Q80Z61	Q80Z61 rattus norv
29	9	1.1	386	16 Q8RE76	Q8RE76 thermoanaer
30	9	1.1	393	16 Q8PMW2	Q8PMW2 xanthomonas
31	9	1.1	393	16 Q8PB63	Q8PB63 xanthomonas
32	9	1.1	418	13 Q7ZYV0	Q7ZYV0 brachydanio
33	9	1.1	481	16 Q92VS2	Q92VS2 rhizobium m
34	9	1.1	612	8 Q94TA9	Q94TA9 neoscapelus
35	9	1.1	681	11 Q8BHP7	Q8BHP7 mus musculu
36	8	1.0	27	2 Q44952	Q44952 borrelia bu
37	8	1.0	34	4 Q96IX0	Q96IX0 homo sapien
38	8	1.0	40	16 Q9IIX6	Q9IIX6 pseudomonas
39	8	1.0	59	5 Q61591	Q61591 ostertagia
40	8	1.0	105	3 Q9P8L0	Q9P8L0 gibberella
41	8	1.0	106	11 Q9JII2	Q9JII2 rattus norv
42	8	1.0	120	5 P90945	P90945 caenorhabdi
43	8	1.0	132	16 Q9RXV0	Q9RXV0 deinococcus
44	8	1.0	136	10 Q7XQY7	Q7XQY7 cryza sativ
45	8	1.0	140	4 Q9NXX2	Q9NXX2 homo sapien

ALIGNMENTS

RESULT 1

Q80VY6	PRELIMINARY;	PRT;	753 AA.
ID	Q80VY6		
AC	Q80VY6		
DT	01-JUN-2003 (TRENBLrel. 24, Created)		
DT	01-JUN-2003 (TRENBLrel. 24, Last sequence update)		
DT	01-OCT-2003 (TRENBLrel. 25, Last annotation update)		
DE	Hypothetical protein (Fragment).		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Embryo;		
RX	MEDLINE=2388257; PubMed=12477932;		
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,		
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,		
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,		
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,		
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,		
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,		
RA	Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,		
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,		
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,		
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,		
RA	Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,		
RT	Jones S.J., Marra M.A.;		
RT	"Generation and initial analysis of more than 15,000 full-length human		
RT	and mouse cDNA sequences.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Embryo;		

RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC052189; AAH52189.1; -;
DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.

DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR00233; Cadherin_C_term.

DR Pfam; PF00028; cadherin; 5.
DR Pfam; PF01049; Cadherin_C_term; 1.

DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 4.

DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 4.

KW Hypothetical protein.
FT NON_TER 1

SQ SEQUENCE 753 AA; 83300 MW; 414AFB4059CE94D2 CRC64;

Query Match 6.2%; Score 51; DB 11; Length 753;

Best Local Similarity 100.0%; Pred. No. 2.7e-43;

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 684 KEPILLPEDTRDNVYFYGEGGEDQDYDTQLHRGLEARPEVVLNDV 734

DB 608 KEPILLPEDTRDNVYFYGEGGEDQDYDTQLHRGLEARPEVVLNDV 658

RESULT 2

Q8BSL6

ID Q8BSL6 PRELIMINARY; PRT; 821 AA.

AC Q8BSL6; 2003 (TREMBLrel. 23, Created)

DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Cadherin 3.

GN CDH3.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J;

RE MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs."

RL Nature 420:563-573(2002).

RL EMBL; AK045041; BAC32194.1; -.

DR MGD; MGI:88356; Cdh3.

DR GO; GO:0005886; C:plasma membrane; IEA.

DR InterPro; IPR002126; Cadherin.

DR InterPro; IPR00233; Cadherin_C_term.

DR Pfam; PF00028; cadherin; 5.

DR Pfam; PF01049; Cadherin_C_term; 1.

DR PRINTS; PR00205; CADHERIN.

DR SMART; SM00112; CA; 4.

DR PROSITE; PS00232; CADHERIN_1; 3.

DR PROSITE; PS0268; CADHERIN_2; 4.

SQ SEQUENCE 821 AA; 90512 MW; 169B56C5FB5CCBD0 CRC64;

Query Match 6.2%; Score 51; DB 11; Length 821;

Best Local Similarity 100.0%; Pred. No. 2.9e-43;

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 684 KEPILLPEDTRDNVYFYGEGGEDQDYDTQLHRGLEARPEVVLNDV 734

DB 676 KEPILLPEDTRDNVYFYGEGGEDQDYDTQLHRGLEARPEVVLNDV 726

RESULT 3

Q8BSL6

ID Q8BSL6 PRELIMINARY; PRT; 822 AA.

AC Q8BSL6; 2003 (TREMBLrel. 23, Created)

DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE Cadherin 3.

GN CDH3.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J;

RE MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs."

RL Nature 420:563-573(2002).

RL EMBL; AK045041; BAC32194.1; -.

DR MGD; MGI:88356; Cdh3.

DR GO; GO:0005886; C:plasma membrane; IEA.

DR InterPro; IPR002126; Cadherin.

DR InterPro; IPR00233; Cadherin_C_term.

DR Pfam; PF00028; cadherin; 5.

DR Pfam; PF01049; Cadherin_C_term; 1.

DR PRINTS; PR00205; CADHERIN.

DR SMART; SM00112; CA; 4.

DR PROSITE; PS00232; CADHERIN_1; 3.

DR PROSITE; PS0268; CADHERIN_2; 4.

SQ SEQUENCE 822 AA; 90611 MW; 7C653D58210A595A CRC64;

Query Match 6.2%; Score 51; DB 11; Length 822;

Best Local Similarity 100.0%; Pred. No. 2.9e-43;

Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 684 KEPILLPEDTRDNVYFYGEGGEDQDYDTQLHRGLEARPEVVLNDV 734

DB 677 KEPILLPEDTRDNVYFYGEGGEDQDYDTQLHRGLEARPEVVLNDV 727

Query Match 4.5%; Score 37; DB 11; Length 105;

Best Local Similarity 100.0%; Pred. No. 2.9e-43;

Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 684 KEPILLPEDTRDNVYFYGEGGEDQDYDTQLHRGLEARPEVVLNDV 734

DB 677 KEPILLPEDTRDNVYFYGEGGEDQDYDTQLHRGLEARPEVVLNDV 727

```

Best Local Similarity 100.0%; Pred. No. 1.5e-29; Indels 0; Gaps 0;
Matches 37; Conservative 0; Mismatches 0;

QY 739 ITPMYRPPANPDIGNFIENLKAANTDTPAPYD 775
    |||||
Db 29 ITPMYRPPANPDIGNFIENLKAANTDTPAPYD 65
    |||||

RESULT 5
Q9UII8 PRELIMINARY; PRT; 821 AA.
AC Q9UII8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE E-cadherin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RA Shibamoto S., Fukudome Y., Yanagihara K.;
RT "Mutant E-cadherin.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC EMBL; AB025105; BAA88956.1; -.
DR HSSP; P09803; 1SUH.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0003509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR Pfam; PF00028; cadherin; 4.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 3.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 3.
DR Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
KW SEQUENCE 821 AA; 90941 MW; 4EFT06672ACE35E9 CRC64;
SQ

Query Match 1.8%; Score 15; DB 4; Length 821;
Best Local Similarity 100.0%; Pred. No. 6.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 RFKLLADMYGGGDD 829
    |||||
Db 807 RFKLLADMYGGGDD 821
    |||||

RESULT 6
Q9UII7 PRELIMINARY; PRT; 901 AA.
AC Q9UII7;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE E-cadherin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RA Shibamoto S., Fukudome Y.;
RT "E-cadherin mutant.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC EMBL; AB025105; BAA88956.1; -.
DR HSSP; P09803; 1SUH.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0003509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR Pfam; PF00028; cadherin; 4.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 3.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 3.
DR Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
KW SEQUENCE 901 AA; 99693 MW; 42160D749BCF120E CRC64;
SQ

Query Match 1.8%; Score 15; DB 4; Length 901;
Best Local Similarity 100.0%; Pred. No. 7.3e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 RFKLLADMYGGGDD 829
    |||||
Db 887 RFKLLADMYGGGDD 901
    |||||

RESULT 7
O18922 PRELIMINARY; PRT; 40 AA.
AC O18922;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Epithelial-cadherin (E-cadherin) (Uvomorulin) (CaM 120/80) (Fragment).
GN CAD1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=RETINA;
RA Lutz D.A., Zheng J.J.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES. E-CADHERIN HAS A POTENT
CC INVASIVE SUPPRESSOR ROLE.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
DR EMBL; AF033019; AAB87474.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR PRINTS; PR00205; CADHERIN.
DR PROSITE; PS00232; CADHERIN_1; PARTIAL.
DR PROSITE; PS0268; CADHERIN_2; PARTIAL.
KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW Calcium-binding; Repeat.
FT NON_TER 1
FT CHAIN <1 >40 EPIITHELIAL-CADHERIN.

```

```

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC EMBL; AB025106; BAA88957.1; -.
DR HSSP; P09803; 1SUH.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR002233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 3.
DR Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
KW SEQUENCE 901 AA; 99693 MW; 42160D749BCF120E CRC64;
SQ

Query Match 1.8%; Score 15; DB 4; Length 901;
Best Local Similarity 100.0%; Pred. No. 7.3e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 RFKLLADMYGGGDD 829
    |||||
Db 887 RFKLLADMYGGGDD 901
    |||||

RESULT 7
O18922 PRELIMINARY; PRT; 40 AA.
AC O18922;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Epithelial-cadherin (E-cadherin) (Uvomorulin) (CaM 120/80) (Fragment).
GN CAD1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=RETINA;
RA Lutz D.A., Zheng J.J.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES. E-CADHERIN HAS A POTENT
CC INVASIVE SUPPRESSOR ROLE.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
DR EMBL; AF033019; AAB87474.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR PRINTS; PR00205; CADHERIN.
DR PROSITE; PS00232; CADHERIN_1; PARTIAL.
DR PROSITE; PS0268; CADHERIN_2; PARTIAL.
KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW Calcium-binding; Repeat.
FT NON_TER 1
FT CHAIN <1 >40 EPIITHELIAL-CADHERIN.

```

FT DOMAIN <1 >40 CYTOPLASMIC (POTENTIAL).
 FT NON TER 40 40
 SQ SEQUENCE 40 AA; 4449 MW; CB928720A51B8372 CRC64;

Query Match 1.7%; Score 14; DB 6; Length 40;
 Best Local Similarity 100.0%; Pred. No. 4.5e-06;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 800 SDQDDYDYLNEWG 813
 Db 25 SDQDDYDYLNEWG 38

RESULT 8
 Q90237 PRELIMINARY; PRT; 864 AA.

AC Q90237; 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE E-cadherin.

GN CDH1.
 OS Brachydanio rerio (zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OC NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21270060; PubMed=11376490;
 RA Babb S.G., Barnett J., Doedens A.L., Cobb N., Liu Q., Sorkin B.C.,
 RA Velick P.C., Raymond P.A., Marrs J.A.;
 RT "Zebrafish E-cadherin: expression during early embryogenesis and
 RT regulation during brain development.";
 RL Dev. Dyn. 221:231-237(2001).

CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
 CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).

DR EMBL; AF364811; AKS2054.1; -.
 DR ZFIN; ZFIN-GENE-010606-1; cdh1.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
 DR InterPro; IPR002126; Cadherin.
 DR InterPro; IPR000233; Cadherin_C term.
 DR Pfam; PF01049; Cadherin_C term; 5.
 DR PRINTS; PR00205; CADHERIN.
 DR SMART; SM00112; CA; 4.
 DR PROSITE; PS00232; CADHERIN_1; 3.
 DR PROSITE; PS0288; CADHERIN_2; 4.
 DR KX Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
 SQ SEQUENCE 864 AA; 95254 MW; 5ACA19538396397C CRC64;

Query Match 1.7%; Score 14; DB 13; Length 864;
 Best Local Similarity 100.0%; Pred. No. 7.6e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 815 RFKKLADMYGGGDD 828
 Db 851 RFKKLADMYGGGDD 864

RESULT 9

Q8MY78 PRELIMINARY; PRT; 2909 AA.

AC Q8MY78; 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Ap-cadherin.
 GN APCAD.
 OS Asterina pectinifera (Starfish).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Asterozoa;
 OC Asteroidea; Valvatacea; Valvatida; Asterinidae; Asterina.
 OC NCBI_TaxID=7594;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Oda H., Wada H., Tagawa K., Akiyama-Oda Y., Satoh N., Humphreys T.,
 RA Zhang S., Tsukita S.;
 RT "A novel amphioxus cadherin that localizes to epithelial adherens
 RT junctions has an unusual domain organization with implications for
 RT chordate phylogeny.";
 RL Evol. Dev. 0:0-0(2002).
 CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
 CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).

DR EMBL; AB075365; BAC06934.1; -.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005509; F:calcium ion binding; IEA.
 DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
 DR InterPro; IPR000152; Asx-Hydroxyl_S.
 DR InterPro; IPR000126; Cadherin.
 DR InterPro; IPR000233; Cadherin_C term.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR006210; IEGF.
 DR InterPro; IPR001791; Laminin_G.
 DR Pfam; PF00028; cadherin; 18.
 DR Pfam; PF01049; Cadherin_C term; 1.
 DR Pfam; PF00008; EGF; 4.
 DR Pfam; PF00054; Laminin G; 2.
 DR PRINTS; PR00205; CADHERIN.
 DR SMART; SM00112; CA; 17.
 DR SMART; SM00181; EGF; 4.
 DR SMART; SM00179; EGF_CA; 4.
 DR SMART; SM00282; LamG; 2.
 DR PROSITE; PS0010; ASX-HYDROXYL; 2.
 DR PROSITE; PS00232; CADHERIN_1; 5.
 DR PROSITE; PS0268; CADHERIN_2; 17.
 DR PROSITE; PS00022; EGF_1; 3.
 DR PROSITE; PS01186; EGF_2; 4.
 DR PROSITE; PS00025; LAM_G DOMAIN; 2.
 DR KX Calcium; Calcium-binding; Cell adhesion; EGF-like domain;
 KW Glycoprotein.
 SQ SEQUENCE 2909 AA; 316542 MW; 2466441EC36619CF CRC64;

Query Match 1.7%; Score 14; DB 5; Length 2909;
 Best Local Similarity 100.0%; Pred. No. C.00023;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 816 FKKLADMYGGGDD 829
 Db 2896 FKKLADMYGGGDD 2909

RESULT 10

Q9H1G6 PRELIMINARY; PRT; 566 AA.

AC Q9H1G6; 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE D1107C24.2 (Continued from BA723K16.1 in Em:AL365229, BA429B11.1 in
 DE Em:AL162457 and BA489M19.1 in Em:AL365401) (Fragment).

GN CH4.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wall M.,
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AL109911; CAC22311.1; -.
DR HSSP; P15116; INCU.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 3.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 3.
DR PROSITE; PS00232; CADHERIN_1; 2.
DR PROSITE; PS0268; CADHERIN_2; 2.
DR Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
KW NON TER 1
FT SEQUENCE 566 AA; 61628 MW; 99A9ADD91B5AE7A3 CRC64;
SQ

Query Match 1.6%; Score 13; DB 4; Length 566;
Best Local Similarity 100.0%; Pred. No. 0.00056;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 RPKKLADMYGGGE 827
Db 552 RPKKLADMYGGGE 564
|||||

RESULT 11
Q9NL47 PRELIMINARY; PRT; 798 AA.
AC Q9NL47
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Cadherin.
GN C1-CADHERIN.
OS Ciona intestinalis.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
OC Phlebobranchia; Cionidae; Ciona.
OX NCBI_TaxID=7719;
RN [1]
RP SEQUENCE FROM N.A.
RA Imai K., Takada N., Satoh N., Satou Y.;
RT "an essential role of beta-catenin in the endoderm specification of
RT ascidian embryo.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AB031540; BAA92182.1; -.
DR HSSP; P15116; INCU.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000233; Cadherin_C_term.

DR Pfam; PF00028; cadherin; 4.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 5.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 5.
KW Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
SQ SEQUENCE 798 AA; 86140 MW; 6F13E42D9F230A5C CRC64;

Query Match 1.6%; Score 13; DB 5; Length 798;
Best Local Similarity 100.0%; Pred. No. 0.00077;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 RPKKLADMYGGGE 827
Db 784 RPKKLADMYGGGE 796
|||||

RESULT 12
Q9NB64 PRELIMINARY; PRT; 824 AA.
ID Q9NB64
AC Q9NB64
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ34177.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Hotta T., Hirakawa S., Murakawa K., Takiguchi S., Kusano J., Chiba Y.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
RA Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuc K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagaatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AK031456; BAC03677.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 5.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 5.
DR PROSITE; PS00232; CADHERIN_1; 3.
DR PROSITE; PS0268; CADHERIN_2; 3.
KW Hypothetical protein; Calcium; Calcium-binding; Cell adhesion;
KW Glycoprotein.
SQ SEQUENCE 824 AA; 90261 MW; 31A0C127BD40BAEB CRC64;

Query Match 1.6%; Score 13; DB 4; Length 824;
Best Local Similarity 100.0%; Pred. No. 0.0008;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 RPKKLADMYGGGE 827
Db 810 RPKKLADMYGGGE 822
|||||

```

RESULT 13
Q80XX1 ID Q80XX1 PRELIMINARY; PRT; 68 AA.
AC Q80XX1;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE N-cadherin (Fragment).
GN N-CADHERIN.
OS Mus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10095;
[1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=92409532; PubMed=1528849;
RA Miyatani S., Copeland N.G., Gilbert D.J., Jenkins N.A., Takeichi M.;
RT "Genomic structure and chromosomal mapping of the mouse N-cadherin
RT gene."
RL Proc. Natl. Acad. Sci. U.S.A. 89:8443-8447(1992).
DR EMBL; S45013; AAP13881.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF01049; Cadherin_C_term; 1.
FT NON_TER 1
SQ SEQUENCE 68 AA; 7130 MW; 860927ADDC8D6595 CRC64;

Query Match 1.4%; Score 12; DB 11; Length 68;
Best Local Similarity 100.0%; Pred. No. 0.0087;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 815 RFKKLADMYGGG 826
Db 55 RFKKLADMYGGG 66

RESULT 14
Q8ND09 ID Q8ND09 PRELIMINARY; PRT; 140 AA.
AC Q8ND09;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN DKFZP434K1130.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN RP SEQUENCE FROM N.A.
RX TISSUE=Testis;
RA Blum H., Bauersachs S., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL834206; CAD38893.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF01049; Cadherin_C_term; 1.
KW Hypothetical protein.
SQ SEQUENCE 140 AA; 15832 MW; C4B81EBFB31F812B CRC64;

Query Match 1.4%; Score 12; DB 4; Length 140;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 703 EGGGEEDQDYD 714
Db 359 EGGGEEDQDYD 370

Query Match 1.4%; Score 12; DB 13; Length 490;
Best Local Similarity 100.0%; Pred. No. 0.0054;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 703 EGGGEEDQDYD 714
Db 359 EGGGEEDQDYD 370

Search completed: September 21, 2004, 22:25:24
Job time : 138 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:09:44 ; Search time 132 Seconds
(without alignments)
1774.484 Million cell updates/sec

Title: US-09-916-849a-1

Perfect score: 829

Sequence: 1 MGLPRGPLASLLILQVCWLQ.....NEWGSRFKKADMYGGGDD 829

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	829	100.0	829	5	ABG61897	Abg61897 Prostate
2	829	100.0	829	5	ABB81476	Abb81476 Human P-C
3	829	100.0	829	5	ABJ05598	Abj05598 Breast ca
4	829	100.0	829	5	AAU97432	AAU97432 Human P-C
5	829	100.0	829	5	AAM50864	AAM50864 Cadherin
6	829	100.0	829	6	ABP58357	ABP58357 Human P-C
7	829	100.0	829	7	ADC15497	ADC15497 Human bas
8	829	100.0	829	7	ADD14190	ADD14190 Human src
9	728	87.8	829	5	ABP54683	ABP54683 Metastati
10	728	87.8	829	6	ABR58670	ABR58670 Human can
11	728	87.8	829	6	ABU56434	ABU56434 Lung can
12	673	81.2	829	6	ABU56670	ABU56670 Lung can
13	108	13.0	108	2	AAU09093	AAU09093 Human P-C
14	108	13.0	108	2	AAU17104	AAU17104 Human P-C
15	108	13.0	108	3	AAU78184	AAU78184 Human P-C
16	108	13.0	108	4	AAG65364	AAG65364 Human P-C
17	108	13.0	108	5	AAM47471	AAM47471 Human P-C
18	108	13.0	108	5	ABU50282	ABU50282 Human P-C
19	108	13.0	108	7	ABO43533	ABO43533 Human P-C
20	108	13.0	108	7	ABW01245	ABW01245 Human cad
21	106	12.8	106	3	AAU78193	AAU78193 Human P-C
22	106	12.8	106	6	ABU60291	ABU60291 Human P-C
23	106	12.8	106	7	ABW01254	ABW01254 Human cad
24	48	5.8	78	2	AAU08239	AAU08239 Human cad
25	43	5.2	43	2	AAU91055	AAU91055 Flanking

26	43	5.2	43	3	AAU27069	Abg27069 Beta-cate
27	22	2.7	108	2	AAU09094	AAU09094 Mouse P-C
28	22	2.7	108	2	AAU17105	AAU17105 Mouse P-C
29	22	2.7	108	3	AAU78185	AAU78185 Mouse P-C
30	22	2.7	108	4	AAU55365	AAU55365 Mouse P-C
31	22	2.7	108	5	AAU47472	AAU47472 Murine P-
32	22	2.7	108	6	ABU60283	ABU60283 Mouse P-C
33	22	2.7	108	7	ABO43534	ABO43534 Mouse P-C
34	22	2.7	108	7	ABW01246	ABW01246 Mouse cad
35	19	2.3	19	5	AAM50867	AAM50867 Cadherin
36	19	2.3	19	7	ADC15500	ADC15500 Human bas
37	17	2.1	17	5	AAM50868	AAM50868 Cadherin
38	17	2.1	17	7	ADC15501	ADC15501 Human bas
39	15	1.8	15	5	AAM50869	AAM50869 Cadherin
40	15	1.8	15	7	ADC15502	ADC15502 Human bas
41	15	1.8	882	2	AAU09375	AAU09375 Wild-type
42	15	1.8	882	3	AAU35730	AAU35730 Human E-C
43	15	1.8	882	4	AAU73490	AAU73490 Human E-C
44	15	1.8	882	5	ABG96293	ABG96293 Human ova
45	15	1.8	882	5	AAU78051	AAU78051 Human E-C

ALIGNMENTS

RESULT 1

ABG61897
ID ABG61897 standard; protein; 829 AA.

XX ABG61897;

XX
DT 15-AUG-2002 (first entry)

XX Prostate cancer-associated protein #98.

XX Prostate cancer; prostate tumour tissue; human; mammal; cytostatic.

XX Mammalia.

XX WO200230268-A2.

XX 18-APR-2002.

XX 12-OCT-2001; 2001WO-US032045.

XX 13-OCT-2000; 2000US-00687576.

XX 08-DEC-2000; 2000US-0073288.

XX 08-DEC-2000; 2000US-00733742.

XX 24-JAN-2001; 2001US-0263957P.

XX 16-MAR-2001; 2001US-0276791P.

XX 16-MAR-2001; 2001US-0276888P.

XX 06-APR-2001; 2001US-0281922P.

XX 24-APR-2001; 2001US-0286214P.

XX 30-APR-2001; 2001US-00847046.

XX 04-MAY-2001; 2001US-0288589P.

XX (EOSB-) EOS BIOTECHNOLOGY INC.

XX Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;

XX WPI, 2002-471335/50.

XX N-PSDB; ABR92214.

XX Detecting a prostate cancer-associated transcript in a cell in a patient,

XX useful for diagnosing prostate cancer (PC) or screening modulators of PC,

XX by determining if prostate cancer-associated genes are expressed in a

XX prostate tissue.

XX Claim 27; Page 384-385; 436pp; English.

The present invention relates to methods of detecting a prostate cancer-associated transcript in a cell from a patient. The method comprises contacting a biological sample from the patient with prostate cancer-

CC associated polynucleotides (designated PC genes) that selectively
 CC hybridize to a sequence that is at least 80% identical to them. The
 CC prostate cancer-associated polynucleotide sequences are differentially
 CC expressed in prostate tumour tissue or in prostate cancer and are derived
 CC from the tissues of various organisms such as humans or other mammals
 CC (e.g. mice, sheep and dogs). The methods of the invention are useful for
 CC diagnosing and treating prostate cancer in mammals. The prostate cancer-
 CC associated genes are useful for diagnosing or treating prostate cancer,
 CC as well as for identifying modulators of prostate cancer or agents that
 CC inhibit prostate cancer. The nucleic acid sequences are particularly
 CC useful in gene therapy, as a vaccine or in antisense applications.
 CC ABG61800-ABG61944 represent prostate cancer-associated proteins
 XX
 XX Sequence 829 AA;

Query Match 100.0%; Score 829; DB 5; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGLASLLQLQVCWLCRAASEPCRAVFAEVTLEAGAEQPGQALGVFMGCPG 60
 DB 1 MGLPRGLASLLQLQVCWLCRAASEPCRAVFAEVTLEAGAEQPGQALGVFMGCPG 60

QY 61 QEPALFSTNDPFTVNGETVQERRSLKERNPLKIPFSKRILRRHKRDVWVAPISVPENG 120
 DB 61 QEPALFSTNDPFTVNGETVQERRSLKERNPLKIPFSKRILRRHKRDVWVAPISVPENG 120

QY 121 KGPFPORLNKLNKRDRTKIFYSITGPGADSPPEGVFAVEKGTGLLANKLPDREETA 180
 DB 121 KGPFPORLNKLNKRDRTKIFYSITGPGADSPPEGVFAVEKGTGLLANKLPDREETA 180

QY 181 YELFCHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTRFGSVLEGVLPGTSVMQVAT 240
 DB 181 YELFCHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTRFGSVLEGVLPGTSVMQVAT 240

QY 241 DEDDAITYNGVAVYAIHQEPKDPHLMFTTHRSSTISVSSGLDREKVPYTLTQA 300
 DB 241 DEDDAITYNGVAVYAIHQEPKDPHLMFTTHRSSTISVSSGLDREKVPYTLTQA 300

QY 301 TMDGSGSTTTAVAVVEILDANDNAPFDPQKYEAVHPENAVGHEVQRLTVTDLPANSP 360
 DB 301 TMDGSGSTTTAVAVVEILDANDNAPFDPQKYEAVHPENAVGHEVQRLTVTDLPANSP 360

QY 361 AWRATYLIMGDDGDHFTTTHPESQGLTTRKGLDPEAKNQHTLYVEVTNEAPFVLK 420
 DB 361 AWRATYLIMGDDGDHFTTTHPESQGLTTRKGLDPEAKNQHTLYVEVTNEAPFVLK 420

QY 421 PTSTATVHVHEDVNEAPVFPVPSKVVEQGIPTGEPVCVVTADDPKKNOKISYRI 480
 DB 421 PTSTATVHVHEDVNEAPVFPVPSKVVEQGIPTGEPVCVVTADDPKKNOKISYRI 480

QY 481 DPAGWLAMPDSCQVAVGLDREDQFVRNNIYEVWVLAMDNQSPPTGTGTLTLTID 540
 DB 481 DPAGWLAMPDSCQVAVGLDREDQFVRNNIYEVWVLAMDNQSPPTGTGTLTLTID 540

QY 541 VNDHGVPPEPQITICNQSPPVRLNITDKLSPTSPFQALTDSDIYTWAEVNEGSD 600
 DB 541 VNDHGVPPEPQITICNQSPPVRLNITDKLSPTSPFQALTDSDIYTWAEVNEGSD 600

QY 601 TVVLSLKLKQDPTDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKGFIPLVL 660
 DB 601 TVVLSLKLKQDPTDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKGFIPLVL 660

QY 661 GAVLALLFLLVLLLVKRRKIKPELLLPEDTDRNPFYEGEGGEDDQDIDTQLHR 720
 DB 661 GAVLALLFLLVLLLVKRRKIKPELLLPEDTDRNPFYEGEGGEDDQDIDTQLHR 720

QY 721 GLEAREPVVLRNDVAPTIPTPMYRPRANPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 DB 721 GLEAREPVVLRNDVAPTIPTPMYRPRANPDEIGNFIENLKAANTDPTAPPYDTLLVF 780

QY 781 DYEGSGSDAASLSLTSSASDQDQDYDLNWSGRFKKLADMYGGEDD 829

DB 781 DYEGSGSDAASLSLTSSASDQDQDYDLNWSGRFKKLADMYGGEDD 829

RESULT 2
 ABB81476
 ID ABB81476 standard; protein; 829 AA.
 AC ABB81476;
 XX
 XX 30-AUG-2002 (first entry)
 DE Human P-cadherin protein SEQ ID NO:50.
 KW Cadherin; alpha-catenin; cancer; beta-catenin binding domain; melanoma;
 KW o-catenin; colon cancer.
 OS Homo sapiens.
 XX
 XX US2002045591-A1.
 PD 18-APR-2002.
 XX
 XX 17-JUL-2001; 2001US-00905983.
 XX
 XX 26-MAY-1998; 98IL-00124650.
 PR 26-MAY-1999; 99US-00318633.
 XX
 XX (GEIG/) GEIGER B.
 PA (BENZ/) BEN-ZE'EV A.
 PA (SADO/) SADOT E.
 XX
 XX Geiger B, Ben-Ze'ev A, Sadot E;
 XX WPI; 2002-499105/53.
 DR N-PSDB; ABN83393.
 XX
 XX New construct encoding soluble cytoplasmic portion of cadherin including
 XX beta catenin binding domain useful in treating cancer associated with
 XX high beta-catenin activity e.g. colon cancer and melanoma.
 XX
 XX Example 3; Page 53-55; 102pp; English.

CC The present invention describes a pharmaceutical composition for treating
 CC cancer associated with abnormally high beta-catenin activity. The
 CC pharmaceutical composition comprises a gene therapy vehicle harbouring a
 CC polynucleotide that contains: (i) a nucleotide sequence encoding a
 CC soluble cytoplasmic portion of a cadherin which lacks a transmembrane
 CC portion and an extracellular portion of the cadherin, and includes a beta
 CC -catenin binding domain; and (b) an upstream promoter for directing
 CC expression of the soluble cytoplasmic portion of the cadherin in a
 CC mammalian cell. Also described is a pharmaceutical composition for
 CC treating cancer associated with abnormally high activity levels of beta-
 CC catenin comprising a gene therapy vehicle harbouring a polynucleotide
 CC that contains: (a) a nucleotide sequence encoding an o-catenin; and (b)
 CC an upstream promoter for directing expression of the o-catenin in a
 CC mammalian cell. The pharmaceutical compositions have cytostatic activity
 CC and can be used in the suppression of beta-catenin-mediated
 CC transactivation. They can be used for treating cancers associated with
 CC abnormally high activity levels of beta-catenin such as colon cancers and
 CC melanomas, by reducing these high activity levels of beta-catenin in
 CC mammalian cells. The present sequence represents human P-cadherin which
 CC is used in the exemplification of the present invention

Sequence 829 AA;

Query Match 100.0%; Score 829; DB 5; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGLASLLQLQVCWLCRAASEPCRAVFAEVTLEAGAEQPGQALGVFMGCPG 60
 DB 1 MGLPRGLASLLQLQVCWLCRAASEPCRAVFAEVTLEAGAEQPGQALGVFMGCPG 60

QY	61	QBPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPFSKILRRHKRDWVAPISVPENG	120
DB	61	QBPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPFSKILRRHKRDWVAPISVPENG	120
QY	121	KGPFQRLNQLKSNKDRDTKIFYSLTGPGADSPPEGVFAVEKGTGMLLNKPLDREI	180
DB	121	KGPFQRLNQLKSNKDRDTKIFYSLTGPGADSPPEGVFAVEKGTGMLLNKPLDREI	180
QY	181	YELFGHVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSLVGLVPGTSMQV	240
DB	181	YELFGHVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSLVGLVPGTSMQV	240
QY	241	DEDDAIYTNQWVAYSIIHSQPKPHDLMFTIHRSTGTISVISGGLDREKVPYTL	300
DB	241	DEDDAIYTNQWVAYSIIHSQPKPHDLMFTIHRSTGTISVISGGLDREKVPYTL	300
QY	301	TDMDGSGSTTAVAVVEILDANDNAPMDPKQYEAHVPENAVGHVEQRLTVTDL	360
DB	301	TDMDGSGSTTAVAVVEILDANDNAPMDPKQYEAHVPENAVGHVEQRLTVTDL	360
QY	361	AWRATYLIMGDDGDGHFTITTHPSNQILTRKGLDPEAKNQHTLYVEVTNEAP	420
DB	361	AWRATYLIMGDDGDGHFTITTHPSNQILTRKGLDPEAKNQHTLYVEVTNEAP	420
QY	421	PTSTATIVVHVEDVNEAPVFPVPPSKVVEVQEGIPTGPPVCVYAEPPDKENQ	480
DB	421	PTSTATIVVHVEDVNEAPVFPVPPSKVVEVQEGIPTGPPVCVYAEPPDKENQ	480
QY	481	DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVVWLAMDNGSPPTGTG	540
DB	481	DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVVWLAMDNGSPPTGTG	540
QY	541	VNDHGPVPEPROITTCNOSPVRHVLIITDKDLSPHTSFFQAQLTDDSDIY	600
DB	541	VNDHGPVPEPROITTCNOSPVRHVLIITDKDLSPHTSFFQAQLTDDSDIY	600
QY	601	TVVLSLKKFLKQDITYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPG	660
DB	601	TVVLSLKKFLKQDITYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPG	660
QY	661	GAVALALLFLLVLLVLLVRKKIKIPEPLLPEDDTRDNVYVYGGEGGEEDQ	720
DB	661	GAVALALLFLLVLLVLLVRKKIKIPEPLLPEDDTRDNVYVYGGEGGEEDQ	720
QY	721	GLEARPEVLRNDVAPTIIPTPMYRPRANPDEIGNFIIENLKAANTDPTAPP	780
DB	721	GLEARPEVLRNDVAPTIIPTPMYRPRANPDEIGNFIIENLKAANTDPTAPP	780
QY	781	DYEGSGSDAASLSLTSASDQDQDYDYLNWGSRFKKLADMYCGGDD	829
DB	781	DYEGSGSDAASLSLTSASDQDQDYDYLNWGSRFKKLADMYCGGDD	829
RESULT 3			
ABJ05598			
XX	ID	ABJ05598 standard; protein; 829 AA.	
XX	AC	ABJ05598;	
XX	DT	14-NOV-2002 (first entry)	
XX	XX	Breast cancer-associated protein 53.	
XX	DE	Breast cancer; breast cancer-associated gene sequence; drug development;	
XX	KW	pharmacogenetics; biosensor development.	
XX	KW	Unidentified.	
XX	OS	WO200259377-A2.	
XX	PN	01-AUG-2002.	
XX	PD	24-JAN-2002; 2002WO-US002242.	
XX	PF		

[illegible]

Sequence 829 AA;

Query Match	100.0%;	Score 829;	DB 5;	Length 829;
Best Local Similarity	100.0%;	Pred. No. 0;		

QY	1	MGLPRGFLASILLLOVCWLQCAASEPCRAVPREAEVTLLEAGGABQEPQALGKVFMGCPG	60
Db	1	MGLPRGFLASILLLOVCWLQCAASEPCRAVPREAEVTLLEAGGABQEPQALGKVFMGCPG	60
QY	61	QEPALFSTDNDDFTVRNGETVQERRSIKERNPLKIFPSKILIRHKDWWVAPISVPENG	120
Db	61	QEPALFSTDNDDFTVRNGETVQERRSIKERNPLKIFPSKILIRHKDWWVAPISVPENG	120
QY	121	KGPPFQRLNQLKSNKDRDKTIFYITGPGADSPPEGVFAVEKETGWLILNKLPLDREEIAK	180
Db	121	KGPPFQRLNQLKSNKDRDKTIFYITGPGADSPPEGVFAVEKETGWLILNKLPLDREEIAK	180
QY	181	YELFGHAVSENGASVEDPMNISIITVTQNDHKPKFTQDTPRGSVLEGLPGTSVMQVTAT	240
Db	181	YELFGHAVSENGASVEDPMNISIITVTQNDHKPKFTQDTPRGSVLEGLPGTSVMQVTAT	240
QY	241	DEDDAIITYNGVWAYSITHSQEPKDPHLMFTIHRSTGTISVISGSLDREKVPXYTLTIOA	300
Db	241	DEDDAIITYNGVWAYSITHSQEPKDPHLMFTIHRSTGTISVISGSLDREKVPXYTLTIOA	300
QY	301	TDMOGDGSTTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGEVORLTLTDDLDAPNSP	360
Db	301	TDMOGDGSTTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGEVORLTLTDDLDAPNSP	360
QY	361	AWRATYILMGDDGDGHFTTITHTPSNOCILTRKGLDPEAKNQHTLYVEVTNEAPFVKL	420
Db	361	AWRATYILMGDDGDGHFTTITHTPSNOCILTRKGLDPEAKNQHTLYVEVTNEAPFVKL	420

RESULT 3

ABJ05598
ID ABJ05598 standard: protein: 829 AA.

XX
AC AB.T05598:

XX DE 14 NOV 20

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

KW breast cancer; breast cancer-associated gene sequence; drug development; pharmacogenetics; biosensor development.

XX
XX Identified

XX

XX
XX

XX

PF 24-JAN-2002; 2002WQ-US002242.

QY 421 PTSTATIVHVEDVNEAPVFPBPSKVVEQGIPTGEPVCVYTAEDPDKENQKISYRIILR 480
 DB 421 PTSTATIVHVEDVNEAPVFPBPSKVVEQGIPTGEPVCVYTAEDPDKENQKISYRIILR 480
 QY 481 DPAGWLAMPDSQGVTAAGTLDREDEQFVRNNIYEVNVLAMONGSPPTTGTTGLLLTLID 540
 DB 481 DPAGWLAMPDSQGVTAAGTLDREDEQFVRNNIYEVNVLAMONGSPPTTGTTGLLLTLID 540
 QY 541 VNHGVPPEPRQITICNQSPVRHLNITDKLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
 DB 541 VNHGVPPEPRQITICNQSPVRHLNITDKLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
 QY 601 TVVLSLKKFLKQDITYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
 DB 601 TVVLSLKKFLKQDITYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
 QY 661 GAVLALLFLLLVLLVLRKKRKEPELLLPEDDTRDNVFFYBEGGGEEDQDYDITQLHR 720
 DB 661 GAVLALLFLLLVLLVLRKKRKEPELLLPEDDTRDNVFFYBEGGGEEDQDYDITQLHR 720
 QY 721 GLEAPBEVLNRNDVAPTIIPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDILLVF 780
 DB 721 GLEAPBEVLNRNDVAPTIIPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDILLVF 780
 QY 781 DYEGSGSDAASLSLSSASDQDQDYDYLNEGSRFKKLADMYGGGEDD 829
 DB 781 DYEGSGSDAASLSLSSASDQDQDYDYLNEGSRFKKLADMYGGGEDD 829
 RESULT 4
 AAU97492
 ID AAU97492 standard; protein; 829 AA.
 AC AAU97492;
 XX
 DT 13-AUG-2002 (first entry)
 XX
 DE Human P-cadherin (placental cadherin) protein.
 KW Human; P-cadherin; placental cadherin; solid surface; vascular graft;
 KW cell-to-cell cohesion; endothelial cell-coated surface;
 KW hydrodynamic shear; shaking container; continuous harvest system;
 KW laminar shear device.
 XX
 OS Homo sapiens.
 XX
 PN WO200231121-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 15-OCT-2001; 2001WO-US032030.
 XX
 PR 13-OCT-2000; 2000US-0241216P.
 PR 27-OCT-2000; 2000US-0243693P.
 PR 11-OCT-2001; 2001US-00975723.
 XX
 PA (UYNE-) UNIV NEW JERSEY MEDICINE & DENTISTRY.
 XX
 PI Nackman G, Foty RA;
 XX
 DR WPI; 2002-435443/46.
 DR N-PSDB; ABK52038.
 XX
 PT Populating solid surface especially biomedical devices with vascular
 PT endothelial cells by increasing cell-to-cell cohesion useful to heal
 PT vascular grafts.
 XX
 PS Disclosure; Page 9-10; 38pp; English.
 XX
 CC The present invention relates to a new method of populating a solid
 CC surface with cells. The method of the invention involves increasing the
 CC cell-to-cell cohesion of the cells. The method is useful for populating a
 CC solid surface with human and nonhuman cells. The invention is also useful

CC for achieving better utilisation of vascular grafts and also in any
 CC system that employs endothelial cell-coated surface susceptible to
 CC hydrodynamic shear, such as shaking container coated with endothelial
 CC cells, continuous harvest systems for harvesting the products of
 CC endothelial cells on solid surface, and laminar shear devices. The
 CC present amino acid sequence represents the human P-cadherin (placental
 CC cadherin) protein of the invention
 XX
 SQ Sequence 829 AA;
 Query Match 100.0%; Score 829; DB 5; Length 829;
 Best Local Similarity 100.0%; Pred. NO. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGLPRGLASLLLLQVLCQAASEPCRAVFAEVTLEAGGAEQBPQALGKVFVGCPC 60
 DB 1 MGLPRGLASLLLLQVLCQAASEPCRAVFAEVTLEAGGAEQBPQALGKVFVGCPC 60
 QY 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKILRRHKEDWVAPISVPENG 120
 DB 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKILRRHKEDWVAPISVPENG 120
 QY 121 KGPPQRLNQLKSKNDRDTXIFYSITGPADSPPEGVFAVEKGTGMLLNKPLDRBIAK 180
 DB 121 KGPPQRLNQLKSKNDRDTXIFYSITGPADSPPEGVFAVEKGTGMLLNKPLDRBIAK 180
 QY 181 YELFGHVAVSENGASVEDPMNISIIIVTDNDHKPKFTQDTFRGSLGVLPGTSVMQVAT 240
 DB 181 YELFGHVAVSENGASVEDPMNISIIIVTDNDHKPKFTQDTFRGSLGVLPGTSVMQVAT 240
 QY 241 DEDDAIYTYNGWAVSISHSQEPKDPHLMFTIHRSTGTISVISGLDRKREKPYETITQA 300
 DB 241 DEDDAIYTYNGWAVSISHSQEPKDPHLMFTIHRSTGTISVISGLDRKREKPYETITQA 300
 QY 301 TDMGDGSGTTTAVAVVEILDANDNAPMFDQKYEAVPENAVGHEVORLTVTDLDPNSP 360
 DB 301 TDMGDGSGTTTAVAVVEILDANDNAPMFDQKYEAVPENAVGHEVORLTVTDLDPNSP 360
 QY 361 AWRATYILMGDDGDGHFTITTHPESNQGLITTRKGLDFAKNQHTLYVEVTNEAPFVLKL 420
 DB 361 AWRATYILMGDDGDGHFTITTHPESNQGLITTRKGLDFAKNQHTLYVEVTNEAPFVLKL 420
 QY 421 PTSTATIVHVEDVNEAPVFPBPSKVVEQGIPTGEPVCVYTAEDPDKENQKISYRIILR 480
 DB 421 PTSTATIVHVEDVNEAPVFPBPSKVVEQGIPTGEPVCVYTAEDPDKENQKISYRIILR 480
 QY 481 DPAGWLAMPDSQGVTAAGTLDREDEQFVRNNIYEVNVLAMONGSPPTTGTTGLLLTLID 540
 DB 481 DPAGWLAMPDSQGVTAAGTLDREDEQFVRNNIYEVNVLAMONGSPPTTGTTGLLLTLID 540
 QY 541 VNHGVPPEPRQITICNQSPVRHLNITDKLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
 DB 541 VNHGVPPEPRQITICNQSPVRHLNITDKLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
 QY 601 TVVLSLKKFLKQDITYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
 DB 601 TVVLSLKKFLKQDITYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
 QY 661 GAVLALLFLLLVLLVLRKKRKEPELLLPEDDTRDNVFFYBEGGGEEDQDYDITQLHR 720
 DB 661 GAVLALLFLLLVLLVLRKKRKEPELLLPEDDTRDNVFFYBEGGGEEDQDYDITQLHR 720
 QY 721 GLEAPBEVLNRNDVAPTIIPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDILLVF 780
 DB 721 GLEAPBEVLNRNDVAPTIIPTPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDILLVF 780
 QY 781 DYEGSGSDAASLSLSSASDQDQDYDYLNEGSRFKKLADMYGGGEDD 829
 DB 781 DYEGSGSDAASLSLSSASDQDQDYDYLNEGSRFKKLADMYGGGEDD 829
 RESULT 5
 AAM50864

ID XX AAM50864 standard; protein; 829 AA.
 AC XX AAM50864;
 DT XX 07-MAY-2002 (first entry)
 DE XX Cadherin 3, basal cell marker.
 KW XX Cadherin 3; P-cadherin; human; basal cell; marker; breast cancer; tumour;
 KW XX diagnosis; prognosis; classification; gene therapy.
 OS XX Homo sapiens.
 XX XX WO200208765-A2.
 PN XX 31-JAN-2002.
 PD XX 26-JUL-2001; 2001WO-US023843.
 XX XX 26-JUL-2000; 2000US-0220967P.
 PR XX (STRD) UNIV STANFORD.
 PA XX (GENO-) APPLIED GENOMICS INC.
 XX XX Botstein D, Brown PO, Perou C, Ross D, Van De Rijn M, Ring B;
 PI XX Seitz R;
 XX XX WPI; 2002-206094/26.
 DR XX
 XX XX Classifying and treating breast tumors associated with the expression and
 PT XX activity of cadherin 3 or P-cadherin, matrix metalloproteinase 14 and/or
 PT XX cadherin EGF LAG seven-pass G-type receptor 2 or EGF-like Domain,
 PT XX Multiple 2.
 XX XX
 PS XX Claim 1; Fig 1A; 741pp; English.
 CC XX
 CC XX The present sequence is that of human cadherin 3 (or P-cadherin). The
 CC XX invention provides new reagents and methods for the management (e.g.
 CC XX detection, classification, provision of diagnostic and prognostic
 CC XX information, treatment, etc.) of breast cancer. cDNA microarray
 CC XX technology was used to identify genes whose expression profile across a
 CC XX large group of tumour samples correlated with that of cytokeratin 5 and
 CC XX cytokeratin 17, markers for basal cells of the normal mammary lactation
 CC XX gland. Tumours that express cytokeratin 5/6 and/or 17 have a poor
 CC XX prognosis relative to tumours overall. The basal marker genes identified
 CC XX included those encoding cadherin 3, matrix metalloproteinase 14 (see
 CC XX AAM50866), and cadherin EGF LAG seven-pass G-type receptor 2 (see
 CC XX AAM50866). Detection of these markers allows classification of a tumour
 CC XX sample as belonging to a basal tumour subclass. By providing reagents
 CC XX that may reliably be used to classify tumours as belonging to a basal
 CC XX subclass, the invention enables a variety of methods for improving
 CC XX therapeutic options for patients with breast cancer and for
 CC XX individualising therapy. Potential therapeutic agents include compounds
 CC XX that modulate breast basal cell marker genes or that modulate
 CC XX polypeptides encoded by these genes. In particular, therapeutic agents
 CC XX include antibodies that specifically bind to the basal marker proteins
 CC XX and which recognise basal cells of normal mammary lactation glands. The
 CC XX antibodies can be used either by themselves or when conjugated to, or
 CC XX delivered with, another molecule such as a toxic compound. The likelihood
 CC XX that cadherin 3 is membrane-bound makes it an attractive candidate for
 CC XX antibody therapeutics
 XX XX
 SQ XX
 Query Match 100.0%; Score 829; DB 5; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGLPRGLASLLLLQVCWLOCAASECRAVFEAEVTLAAGAEQEPGOALGKVFNGCPG 60
 DB 1 MGLPRGLASLLLLQVCWLOCAASECRAVFEAEVTLAAGAEQEPGOALGKVFNGCPG 60
 QY 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKRILLRHRKRDWVAPISVPENG 120

Db 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKRILLRHRKRDWVAPISVPENG 120
 QY 121 KGFPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLKLLNKPDLREIAK 180
 Db 121 KGFPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLKLLNKPDLREIAK 180
 QY 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240
 Db 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240
 QY 241 DEDDAIYTVNGVAVSIHSGQPKDHDLMFTIHRSTGTISVISSSGLDRKVPYTTITQA 300
 Db 241 DEDDAIYTVNGVAVSIHSGQPKDHDLMFTIHRSTGTISVISSSGLDRKVPYTTITQA 300
 QY 301 TDMGDSGTTTAVAVVEILDANDNAPFDPQKYEAVPENAVGHEVORLTVTDLDPNSP 360
 Db 301 TDMGDSGTTTAVAVVEILDANDNAPFDPQKYEAVPENAVGHEVORLTVTDLDPNSP 360
 QY 361 AWRATYLINGGDDGDHFTIITHPESNQGLITTRKGLDFAKNQHTLYVEVTNEAPFVKL 420
 Db 361 AWRATYLINGGDDGDHFTIITHPESNQGLITTRKGLDFAKNQHTLYVEVTNEAPFVKL 420
 QY 421 PTSTATIVHVEDVNEAPVFPSPKVEVQEGIPTEGVCVYTAEDPDKENOKISYRILR 480
 Db 421 PTSTATIVHVEDVNEAPVFPSPKVEVQEGIPTEGVCVYTAEDPDKENOKISYRILR 480
 QY 481 DPAGWLAMDPSGOVTAVGTLDRDEQFVRNNIYEVWVLANDNGSPPTTGTLLTLID 540
 Db 481 DPAGWLAMDPSGOVTAVGTLDRDEQFVRNNIYEVWVLANDNGSPPTTGTLLTLID 540
 QY 541 VNDHGVPPEPRQITICNOSPVRHLNITDKLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
 Db 541 VNDHGVPPEPRQITICNOSPVRHLNITDKLSPHTSPFOAQLTDDSDIYWTAEVNEEGD 600
 QY 601 TVVLSLKKFLKQDQYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFLLPVL 660
 Db 601 TVVLSLKKFLKQDQYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFLLPVL 660
 QY 661 GAVLALLFLLLVLLVLLVLRKKRKEIPELPLLEDTRDNVYFYEGEGGEDDQYDITQLHR 720
 Db 661 GAVLALLFLLLVLLVLLVLRKKRKEIPELPLLEDTRDNVYFYEGEGGEDDQYDITQLHR 720
 QY 721 GLEARPEVLRNDVAPITIIPTMYRPPANPDEIGNFIENLKAANTDPTAPPYDILLVF 780
 Db 721 GLEARPEVLRNDVAPITIIPTMYRPPANPDEIGNFIENLKAANTDPTAPPYDILLVF 780
 QY 781 DYEGSGDAASLSLTSASDQDQDYDLNWSGRFKKLADMYGGGEDD 829
 Db 781 DYEGSGDAASLSLTSASDQDQDYDLNWSGRFKKLADMYGGGEDD 829
 RESULT 6
 ABP58357
 ID ABP58357 standard; protein; 829 AA.
 XX AC ABP58357;
 XX DT 07-APR-2003 (first entry)
 XX DE Human P-cadherin.
 XX KW P-cadherin; human; colon cancer; colorectal cancer; cytostatic;
 KW gene therapy.
 OS XX Homo sapiens.
 XX PN WO200297395-A2.
 XX PD 05-DEC-2002.
 XX PF 31-MAY-2002; 2002WO-US017109.
 XX XX

PR 31-MAY-2001; 2001US-0294225P.
 XX (CHIR) CHIRON CORP.
 PA Reinhard C, Klinger J, Jefferson AB, Escobedo J, Randazo F;
 PI Winter J, Goodson R;
 XX N-PSDB; ABZ24736.
 XX WPI; 2003-140501/13.
 XX DR N-PSDB; ABZ24736.
 XX
 PT Inhibiting migration and proliferation of P-cadherin expressing cancer
 PT for treating cancer, especially digestive cancer, characterized by
 PT overexpression of P-cadherin, involves administering a P-cadherin
 PT antagonist.
 XX
 PS Disclosure; Page 10-11; 129pp; English.
 XX
 CC The present sequence is the protein sequence for human P-cadherin
 CC (placental cadherin), a calcium-dependent cellular adhesion protein. The
 CC invention provides methods of treating or diagnosing cancers involving P-
 CC cadherin expression using ligands that target P-cadherin, especially
 CC human anti-P-cadherin antibodies. A claimed method of treating a cancer
 CC characterised by the overexpression and/or upregulation of P-cadherin
 CC comprises the administration of a P-cadherin antagonist, optionally
 CC conjugated to a therapeutic agent. The migration, adhesion and/or
 CC proliferation of the cancer is inhibited, and the method is especially
 CC useful for treating or preventing a digestive cancer such as colon or
 CC colorectal cancer. The antagonists may be an anti-P-cadherin antibody or
 CC its fragment, a ribozyme or antisense oligonucleotide. A transgenic
 CC animal that expresses a recombinant (human) antibody that specifically
 CC binds P-cadherin is also claimed. A method of determining the presence of
 CC cancer involves determining the expression level of P-cadherin in a cell
 CC sample. Also provided are screens for identifying anti-P-cadherin
 CC antibodies have therapeutic activity
 XX
 XX Sequence 829 AA;
 Query Match 100.0%; Score 829; DB 6; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGLPRGLASLLLLQVCWLCQAASEPCRAVFRAEVTLEAGGAEPGQALGVFGCPG 60
 DB 1 MGLPRGLASLLLLQVCWLCQAASEPCRAVFRAEVTLEAGGAEPGQALGVFGCPG 60
 QY 61 QEPALFSTNDFTVRNGETVQERRSLKERNPLKIPPSKRIILRRKRDWVAPISVPENG 120
 DB 61 QEPALFSTNDFTVRNGETVQERRSLKERNPLKIPPSKRIILRRKRDWVAPISVPENG 120
 QY 121 KGFPQRLNQLSKNRDRTKIFYSITPGADSPFGVFAVEXETGWLINLKPDLREEIAK 180
 DB 121 KGFPQRLNQLSKNRDRTKIFYSITPGADSPFGVFAVEXETGWLINLKPDLREEIAK 180
 QY 181 YELFGHAVENGASVEDPWNISIIIVTDQNDHKPKTQDTFRGSVLEGLVPGTSVMQVAT 240
 DB 181 YELFGHAVENGASVEDPWNISIIIVTDQNDHKPKTQDTFRGSVLEGLVPGTSVMQVAT 240
 QY 241 DEDDAIYTYNGVAVYSIHQSEPKDPHLMFTTHRSTGTISVISSGLDREKVPYTLTQA 300
 DB 241 DEDDAIYTYNGVAVYSIHQSEPKDPHLMFTTHRSTGTISVISSGLDREKVPYTLTQA 300
 QY 301 TDMGDGSGTTTAVAVVEILLDANDNAPMDPQKYEAHVPENAVGHEVQRLTVDLDAPNSP 360
 DB 301 TDMGDGSGTTTAVAVVEILLDANDNAPMDPQKYEAHVPENAVGHEVQRLTVDLDAPNSP 360
 QY 361 AMRATYLMGGDDGDHFTTTHPESNQGLITTRKGLDPEAKNQHTLYVEVNEAPFVLKL 420
 DB 361 AMRATYLMGGDDGDHFTTTHPESNQGLITTRKGLDPEAKNQHTLYVEVNEAPFVLKL 420
 QY 421 PTSTATIVVHVEDVNEAPVFPVPPSKVVEQIGPTGEPVCVYTABDDPKENQKISYILR 480
 DB 421 PTSTATIVVHVEDVNEAPVFPVPPSKVVEQIGPTGEPVCVYTABDDPKENQKISYILR 480

QY 481 DPAGWLAMPDPSGGQVTAAGTLDREDEQFVRNNIYEVWVLMANDNGSPPTTGTGTLTLLID 540
 DB 481 DPAGWLAMPDPSGGQVTAAGTLDREDEQFVRNNIYEVWVLMANDNGSPPTTGTGTLTLLID 540
 QY 541 VNDHGPVPEPRQIITICNQSPVRHVNLITDKDLSHTSPFOAQLTDDSDIYWTAEVNEGD 600
 DB 541 VNDHGPVPEPRQIITICNQSPVRHVNLITDKDLSHTSPFOAQLTDDSDIYWTAEVNEGD 600
 QY 601 TWLSLKKFLKQDQYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL 660
 DB 601 TWLSLKKFLKQDQYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPKWGGFILPVL 660
 QY 661 GAVIALFLLLVLLVLRKKIKPEPLLPEDDTRDNVYFYGEGGEEPDQYDITQLHR 720
 DB 661 GAVIALFLLLVLLVLRKKIKPEPLLPEDDTRDNVYFYGEGGEEPDQYDITQLHR 720
 QY 721 GLEARPEVLENDVAPTIIPPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 DB 721 GLEARPEVLENDVAPTIIPPMYRPRPANDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
 QY 781 DYEGSGSDAASLSLTSASDQDQDYDLNWSRPFKFLADMYGGGDD 829
 DB 781 DYEGSGSDAASLSLTSASDQDQDYDLNWSRPFKFLADMYGGGDD 829
 RESULT 7
 ADC15497
 ID ADC15497 standard; protein; 829 AA.
 XX
 AC ADC15497;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Human basal cell marker Cadherin-3.
 XX
 KW Human; basal cell marker; cadherin-3; tumour; P-cadherin;
 KW matrix metalloproteinase 14;
 KW cadherin EGF LAG seven-pass G-type receptor 2;
 KW EGF-like domain Multiple 2; mammary lactation gland; breast tumour;
 KW cancer; cytostatic.
 XX
 OS Homo sapiens.
 XX
 PN US2003086934-A1.
 XX
 PD 08-MAY-2003.
 XX
 PF 26-JUL-2001; 2001US-00918849.
 XX
 PR 26-JUL-2000; 2000US-0220967P.
 XX
 PA (BOTS/) BOTSTEIN D.
 PA (BROW/) BROWN P O.
 PA (PERO/) PEROU C M.
 PA (RING/) RING B.
 PA (ROSS/) ROSS D.
 PA (SEIT/) SEITZ R.
 PA (VRIJ/) VAN DE RIJN J M.
 XX
 PI Borstein D, Brown PO, Perou CM, Ring B, Ross D, Seitz R;
 PI Van De Rijn JM;
 XX
 DR WPI; 2003-786931/74.
 XX
 PT Classifying breast tumor by detecting expression or activity of gene
 PT encoding cadherin 3, matrix metalloproteinase 14, cadherin EGF LAG seven-
 PT pass G-type receptor 2 in tumor sample.
 XX
 PS Claim 1; SEQ ID NO 1; 89pp; English.
 XX
 CC The invention relates to classifying (M1) a tumour comprising detecting
 CC expression or activity of a gene encoding cadherin 3 or P-cadherin
 CC ,matrix metalloproteinase 14 , or cadherin EGF LAG seven-pass G-type

CC receptor 2 or EGF-like domain, Multiple 2 (P1-3) or at least two genes
 CC chosen from gene encoding (P1), (P2) or (P3) in the tumour sample, and
 CC classifying tumour as belonging to a tumour subclass based on results of
 CC detecting step. Also included are testing a subject for expression of P1-
 CC P3 and providing diagnostic, prognostic, or predictive information based
 CC on the detecting step, stratifying the subject for a clinical trial based
 CC on the detecting step, or selecting a treatment based on the detecting
 CC step, an antibody that specifically binds to an epitope found in (P1),
 CC (P2) or (P3) (where the antibody recognises basal cells in normal mammary
 CC lactation glands), methods of testing a compound or a combination of
 CC compounds for activity against tumours. The method (M1) is useful for
 CC classifying a tumour (breast tumour), as belonging to a tumour subclass
 CC such as basal tumour subclass. The antibody is useful for treating a
 CC subject in need of treatment for cancer. The tumour is a breast tumour,
 CC and where the method further involves identifying the tumour as belonging
 CC to a basal tumour subclass. The antibody is conjugated with a toxic
 CC molecule. The present sequence is human Cadherin-3.
 XX
 SQ Sequence 829 AA;
 Query Match 100.0%; Score 829; DB 7; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGLPRGLASLLLLQVCWLCQAASEPCRAVFREAEVTLAEAGAEQPGQALGKVFMGCPG 60
 DB 1 MGLPRGLASLLLLQVCWLCQAASEPCRAVFREAEVTLAEAGAEQPGQALGKVFMGCPG 60
 QY 61 QEPALFTNDNFTVNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVWAPIVPENG 120
 DB 61 QEPALFTNDNFTVNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVWAPIVPENG 120
 QY 121 KGFPPQRLNQLKSKNDRTKIFYSITGPGADSPPEGVFAVEKETGMLLNLKPLDREEIAK 180
 DB 121 KGFPPQRLNQLKSKNDRTKIFYSITGPGADSPPEGVFAVEKETGMLLNLKPLDREEIAK 180
 QY 181 YELFGHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTRFSGVLEGLVPGVSVMQVAT 240
 DB 181 YELFGHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTRFSGVLEGLVPGVSVMQVAT 240
 QY 241 DEDDAITYNGVAVYHSGQEPKDPHLMFTIHRSTGTISVSSGLDREKVPYTLTIOA 300
 DB 241 DEDDAITYNGVAVYHSGQEPKDPHLMFTIHRSTGTISVSSGLDREKVPYTLTIOA 300
 QY 301 TMDGSGSTTAVAVVEILDANDNAPMFPDQKYEAVHPENAVGHEVQRLTVTDLDPNSP 360
 DB 301 TMDGSGSTTAVAVVEILDANDNAPMFPDQKYEAVHPENAVGHEVQRLTVTDLDPNSP 360
 QY 361 AWRATYLMGDDGDHFTTTHPESNQGLTTRKGLDFAKQHTLYVETNEAPVFLKL 420
 DB 361 AWRATYLMGDDGDHFTTTHPESNQGLTTRKGLDFAKQHTLYVETNEAPVFLKL 420
 QY 421 PTSTATIVHVEDNEAPVFPVPSKVVEQGIPTGEPVCVYTAEPDCKENKISYRILR 480
 DB 421 PTSTATIVHVEDNEAPVFPVPSKVVEQGIPTGEPVCVYTAEPDCKENKISYRILR 480
 QY 481 DPAGWLAMPDQGVAVGLDREDQFVRNNIYEVWVLAMDNQSPPTGTGTLTLTLD 540
 DB 481 DPAGWLAMPDQGVAVGLDREDQFVRNNIYEVWVLAMDNQSPPTGTGTLTLTLD 540
 QY 541 VNDHGPVPEPRQITCNQSPVRHLNITDKLSPHTSPFOAQLTDDSDIYTAEVNEED 600
 DB 541 VNDHGPVPEPRQITCNQSPVRHLNITDKLSPHTSPFOAQLTDDSDIYTAEVNEED 600
 QY 601 TVVLSLKFTKQDTYVHLSLSDHGNKEQLTVIRATVCDCHGHEVCPEGKGGFLPVL 660
 DB 601 TVVLSLKFTKQDTYVHLSLSDHGNKEQLTVIRATVCDCHGHEVCPEGKGGFLPVL 660
 QY 661 GAVLALLFLVLLVLLVLRKKRKEPILLPDDTRDNVFFYEGGEGEEDQDIDITQLHR 720
 DB 661 GAVLALLFLVLLVLLVLRKKRKEPILLPDDTRDNVFFYEGGEGEEDQDIDITQLHR 720
 QY 721 GLEARPEVLRNDVAPFIIPTPMYRPRPANDEIGNFIIENLKAANTDPTAPPYDTLLVF 780

DB 721 GLEARPEVLRNDVAPFIIPTPMYRPRPANDEIGNFIIENLKAANTDPTAPPYDTLLVF 780
 QY 781 DYEGSGDAASLSLTSSASDQDDYDLNEMSGRFRKXKLADMYGGGDEDD 829
 DB 781 DYEGSGDAASLSLTSSASDQDDYDLNEMSGRFRKXKLADMYGGGDEDD 829
 RESULT 8
 ADD14190
 ID ADD14190 standard; protein; 829 AA.
 XX
 AC ADD14190;
 DT 01-JAN-2004 (first entry)
 XX
 DE Human src biomarker polypeptide SEQ ID NO:379.
 KW predictor set; protein tyrosine kinase activity modulator;
 KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;
 KW gene therapy; drug sensitivity; genetic profile; cancer; human.
 OS Homo sapiens.
 XX
 PN WO2003062395-A2.
 XX
 PD 31-JUL-2003.
 XX
 PF 17-JAN-2003; 2003WO-US001981.
 XX
 PR 18-JAN-2002; 2002US-0350061P.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Huang F, Fairchild CR, Lee FY, Shaw P;
 XX
 DR WPI; 2003-636735/60.
 DR N-PSDB; ADD14795.
 XX
 PT New polynucleotides and polypeptides for predicting the activity of
 PT compounds that interact with protein tyrosine kinases and/or protein
 PT tyrosine kinase pathways.
 XX
 PS Claim 10; SEQ ID NO 379; 139pp; English.
 XX
 CC The present invention describes a predictor set comprising a plurality of
 CC polynucleotides or polypeptides whose expression pattern is predictive of
 CC the response of cells to treatment with a compound that modulates protein
 CC tyrosine kinase activity or members of the protein tyrosine kinase
 CC pathway. Also described: (1) predicting whether a compound is capable of
 CC modulating the activity of cells, comprising obtaining a sample of cells,
 CC determining whether the cells express a plurality of markers, and
 CC correlating the expression of the markers to the compound's ability to
 CC modulate the activity of the cells; (2) a plurality of cell lines for
 CC identifying polynucleotides and polypeptides whose expression levels
 CC correlate with compound sensitivity or resistance of cells associated
 CC with a disease state; and (3) identifying polynucleotides and
 CC polypeptides that predict compound sensitivity or resistance of cells
 CC associated with a disease state, comprising subjecting the plurality of
 CC cell lines to one or more compounds, analysing the expression pattern of
 CC a microarray of polynucleotides or polypeptides, and selecting
 CC polynucleotides or polypeptides that predict the sensitivity or
 CC resistance of cells associated with a disease state by using the
 CC expression pattern of the microarray. The polynucleotides and
 CC polypeptides have cytostatic activities, and can be used in gene therapy.
 CC The polynucleotides and polypeptides are useful in predicting the
 CC activity of compounds that interact with protein tyrosine kinases and/or
 CC protein tyrosine kinase pathways. These may be used in determining drug
 CC sensitivity in patients to allow the development of individualized
 CC genetic profiles which aid in treating diseases and disorders (e.g.
 CC cancer) based on patient response at a molecular level. The present
 CC sequence is used in the exemplification of the present invention.
 XX

SQ Sequence 829 AA;

Query Match 100.0%; Score 829; DB 7; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASILLQVLCWLCQAASEPCRAVFRAEAVTLEAGGAEQEPGQALGVFMGCPG 60
 DB 1 MGLPRGPLASILLQVLCWLCQAASEPCRAVFRAEAVTLEAGGAEQEPGQALGVFMGCPG 60

QY 61 QEPALFSTDNDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAIPISVPENG 120
 DB 61 QEPALFSTDNDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAIPISVPENG 120

QY 121 KGPFPQRLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREBIAK 180
 DB 121 KGPFPQRLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREBIAK 180

QY 181 YELFGHVAENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240
 DB 181 YELFGHVAENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240

QY 241 DEDDAIYTVNGVVAISHSQEPKDPHDLMTIHRSTGTSIVISSGLDREKVPYTLTIOA 300
 DB 241 DEDDAIYTVNGVVAISHSQEPKDPHDLMTIHRSTGTSIVISSGLDREKVPYTLTIOA 300

QY 301 TMDGSGSTTTAVAVVEILDANDNAPMFDQPKYEAHVPENAVGHVEQRLTVDLDAPNSP 360
 DB 301 TMDGSGSTTTAVAVVEILDANDNAPMFDQPKYEAHVPENAVGHVEQRLTVDLDAPNSP 360

QY 361 AWRATYILMGDDGDHFTIITHPESNOGILTTTRKGLDFAKNQHTLYVEVTEAPVVLK 420
 DB 361 AWRATYILMGDDGDHFTIITHPESNOGILTTTRKGLDFAKNQHTLYVEVTEAPVVLK 420

QY 421 PTSTATIIVHVEDVNEAPVFPVSKVEVOEGIPGTCPEVCVYTAEDPKENOKISVRIIL 480
 DB 421 PTSTATIIVHVEDVNEAPVFPVSKVEVOEGIPGTCPEVCVYTAEDPKENOKISVRIIL 480

QY 481 DPAGWLAMPDSQGVTAAGTLDRDEDFQVRNNIYEVMLAMDNQSPPTTGTGLTLTLD 540
 DB 481 DPAGWLAMPDSQGVTAAGTLDRDEDFQVRNNIYEVMLAMDNQSPPTTGTGLTLTLD 540

QY 541 VNDHGVPEPQITICNOSPVRHVNITDKLSPHSTSPQAQLTDDSDIYWTAEVNEEGD 600
 DB 541 VNDHGVPEPQITICNOSPVRHVNITDKLSPHSTSPQAQLTDDSDIYWTAEVNEEGD 600

QY 601 TVVLSLKKFLKQDYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGGFILPVL 660
 DB 601 TVVLSLKKFLKQDYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGGFILPVL 660

QY 661 GAVLALLFLLVLLVLRKKRKIKEPLLIPEDDTRDNVFIYEGEGGEDQDYDTQLHR 720
 DB 661 GAVLALLFLLVLLVLRKKRKIKEPLLIPEDDTRDNVFIYEGEGGEDQDYDTQLHR 720

QY 721 GLEAREVLRNDVATPIIPTMYRPRRNPBEIGNFIENKAANTDPTAPPYDTLLVF 780
 DB 721 GLEAREVLRNDVATPIIPTMYRPRRNPBEIGNFIENKAANTDPTAPPYDTLLVF 780

QY 781 DYEGSGDAASLSSTSSASDQDDYDYLNEWGRFPKLADMYGGGEDD 829
 DB 781 DYEGSGDAASLSSTSSASDQDDYDYLNEWGRFPKLADMYGGGEDD 829

RESULT 9
 ABP54683
 ID ABP54683 standard; protein; 829 AA.
 XX
 AC ABP54683;
 XX
 DT 30-DEC-2002 (first entry)
 XX
 DE Metastatic colorectal cancer-associated polypeptide.
 XX

KW Colorectal cancer; metastasis; differential expression; cytostatic;
 XX diagnosis; gene therapy; vaccine.
 OS Homo sapiens.
 XX WO200268677-A2.
 PN 06-SEP-2002.
 XX 27-FEB-2002; 2002WO-US006001.
 PF 27-FEB-2001; 2001US-0272206P.
 PR 02-APR-2001; 2001US-0281149P.
 PR 17-APR-2001; 2001US-0284555P.
 XX (EOSB-) EOS BIOTECHNOLOGY INC.
 PA (UYCA-) UNIV CASE WESTERN RESERVE.
 XX Mack DH, Markowitz SD;
 PI WPI; 2002-698677/75.
 DR N-PSDB; ABQ81547.
 XX New genes that are up- or down-regulated in colorectal cancer, useful for
 PT diagnosing colorectal cancer in a subject, or for identifying modulators
 PT of colorectal cancer-associated proteins and genes for treating
 PT colorectal cancer.
 XX Claim 8; Page 245; 260pp; English.
 XX The present sequence is the protein sequence of a human polypeptide
 CC encoded by a gene that exhibits increased expression in colon cancer-
 CC derived metastases compared to normal colon tissue. It is an example of
 CC claimed polypeptides that are encoded by genes which are differentially
 CC expressed in metastatic colorectal cancer cells. Such polypeptides are
 CC useful in diagnostic and prognostic assays, for raising antibodies useful
 CC e.g. in immunotherapy, and in screening for modulator compounds of
 CC therapeutic value
 XX Sequence 829 AA;

Query Match 87.8%; Score 728; DB 5; Length 829;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGLPRGPLASILLQVLCWLCQAASEPCRAVFRAEAVTLEAGGAEQEPGQALGVFMGCPG 60
 DB 1 MGLPRGPLASILLQVLCWLCQAASEPCRAVFRAEAVTLEAGGAEQEPGQALGVFMGCPG 60

QY 61 QEPALFSTDNDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAIPISVPENG 120
 DB 61 QEPALFSTDNDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAIPISVPENG 120

QY 121 KGPFPQRLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREBIAK 180
 DB 121 KGPFPQRLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKETGMLLNKPLDREBIAK 180

QY 181 YELFGHVAENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240
 DB 181 YELFGHVAENGASVEDPMNISIIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240

QY 241 DEDDAIYTVNGVVAISHSQEPKDPHDLMTIHRSTGTSIVISSGLDREKVPYTLTIOA 300
 DB 241 DEDDAIYTVNGVVAISHSQEPKDPHDLMTIHRSTGTSIVISSGLDREKVPYTLTIOA 300

QY 301 TMDGSGSTTTAVAVVEILDANDNAPMFDQPKYEAHVPENAVGHVEQRLTVDLDAPNSP 360
 DB 301 TMDGSGSTTTAVAVVEILDANDNAPMFDQPKYEAHVPENAVGHVEQRLTVDLDAPNSP 360

QY 361 AWRATYILMGDDGDHFTIITHPESNOGILTTTRKGLDFAKNQHTLYVEVTEAPVVLK 420
 DB 361 AWRATYILMGDDGDHFTIITHPESNOGILTTTRKGLDFAKNQHTLYVEVTEAPVVLK 420

QY 421 PTSTATIVVHVEDVNEAPVFPVPPSKVVEQEGIPGTGEPVVCVVTABDPKQKISYRILR 480
 Db 421 PTSTATIVVHVEDVNEAPVFPVPPSKVVEQEGIPGTGEPVVCVVTABDPKQKISYRILR 480
 QY 481 DPAGWLAMPDPSGQVAVGTGLDREDEQFVRNNIYEVWVLMNDGSPPTTGTGTLTLLID 540
 Db 481 DPAGWLAMPDPSGQVAVGTGLDREDEQFVRNNIYEVWVLMNDGSPPTTGTGTLTLLID 540
 QY 541 VNDHGVPPEPRQITICNQSPVRHVNITDKDLSPTSPFQALTDSDIYVTAENVNEGD 600
 Db 541 VNDHGVPPEPRQITICNQSPVRHVNITDKDLSPTSPFQALTDSDIYVTAENVNEGD 600
 QY 601 TVVLSLKXFLKQDITYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGQFIPVL 660
 Db 601 TVVLSLKXFLKQDITYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGQFIPVL 660
 QY 661 GAVLALLFLLVLLVLLVLRKKRKKIKFLLPDDTRDNVFFYGGEGGEDQYDITQLHR 720
 Db 661 GAVLALLFLLVLLVLLVLRKKRKKIKFLLPDDTRDNVFFYGGEGGEDQYDITQLHR 720
 QY 721 GLEARPEVVRNDVAPTITPTMVRPRANPDEIGNFIENLKAANTDPTAPPYDITLIVF 780
 Db 721 GLEARPEVVRNDVAPTITPTMVRPRANPDEIGNFIENLKAANTDPTAPPYDITLIVF 780
 QY 781 DYEGSGDAASLSSTSSASDQDQDYDYLNWGSRFKKLADMYGGGEDD 829
 Db 781 DYEGSGDAASLSSTSSASDQDQDYDYLNWGSRFKKLADMYGGGEDD 829

RESULT 10
 ID ABR58670 standard; protein; 829 AA.
 AC ABR58670;
 DT 09-JUL-2003 (first entry)
 XX Human cancer related protein SEQ ID NO:327.
 DE Human; cancer; diagnosis; screening; modulator; leukaemia; ischaemia;
 KW heart disease; atherosclerosis; endometriosis.
 XX Homo sapiens.
 XX WO2003025138-A2.
 PN 27-MAR-2003.
 PD 17-SEP-2002; 2002WO-US029560.
 PF 17-SEP-2001; 2001US-0323469P.
 PR 20-SEP-2001; 2001US-0323887P.
 PR 13-NOV-2001; 2001US-0350666P.
 PR 08-FEB-2002; 2002US-0355145P.
 PR 08-FEB-2002; 2002US-0355257P.
 PR 12-APR-2002; 2002US-0372246P.
 XX (E0SB-) EOS BIOTECHNOLOGY INC.
 PA Afar D, Aziz N, Gish KC, Hevezi PA, Mack DH, Wilson KE;
 PI Zlotnik A;
 PI WPI: 2003-354600/33.
 DR N-PSDB; ACC72821.

XX New genes that are up-regulated or down-regulated in cancers, useful as
 PT markers for diagnosing e.g. cancer, ischemia or heart diseases, or as
 PT therapeutic targets for screening drugs for treating these diseases.

XX Claim 12; Page 757; 767pp; English.

XX The present invention describes an isolated nucleic acid molecule, which
 CC comprises the sequence of any of the genes that are up-regulated or down-

CC regulated in specific cancers (e.g. about 1031 genes up-regulated in
 CC acute lymphocytic leukemia). ACC72641 to ACC72860 represent cancer
 CC related gene nucleotide sequences which encode the proteins given in
 CC ABR58521 to ABR58709. Also described: (1) determining the presence or
 CC absence of a pathological cell in a patient; (2) an expression vector
 CC comprising a nucleic acid molecule described above; (3) a host cell
 CC comprising the vector; (4) an isolated polypeptide, which is encoded by
 CC the nucleic acid; (5) an antibody that specifically binds the polypeptide
 CC of (4); (6) specifically targeting a compound to a pathological cell in a
 CC patient by administering to the patient the antibody above; and (7) a
 CC drug screening assay. The nucleic acid is useful as diagnostic markers or
 CC therapeutic targets. In particular, the nucleic acid is useful for
 CC diagnosing a pathology, e.g. cancer (e.g. cancer of the bone marrow,
 CC bladder, brain, breast, cervix, colon/rectum, kidney, lung, ovary,
 CC pancreas, prostate, skin and uterus), wounds, ischaemia, heart diseases,
 CC atherosclerosis and endometriosis. The nucleic acid is also useful in
 CC drug screening, particularly for identifying agents for treating these
 CC pathologies

XX SQ Sequence 829 AA;

Query Match 87.8%; Score 728; DB 6; Length 829;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGUPRGPASLILLQVCWLQCAASEPCRAVFAEVTLEAGAEQFPQALGVFGCPG 60
 Db 1 MGUPRGPASLILLQVCWLQCAASEPCRAVFAEVTLEAGAEQFPQALGVFGCPG 60
 QY 61 QEPALFSTDNDDFTVRNGETVQERRSIKERNPLKIPSKRILRRHKRDVVAIPSPENG 120
 Db 61 QEPALFSTDNDDFTVRNGETVQERRSIKERNPLKIPSKRILRRHKRDVVAIPSPENG 120
 QY 121 KGPPPRQLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKGTGMLLNKPLDRSEIAK 180
 Db 121 KGPPPRQLNQLKSNKORDTKIFYSITGPGADSPPEGVFAVEKGTGMLLNKPLDRSEIAK 180
 QY 181 YELFPGHAVSENGASVEDPMNISIIVTQNDHKPKFTQDTPRGVGLGVLPGTSVMQVAT 240
 Db 181 YELFPGHAVSENGASVEDPMNISIIVTQNDHKPKFTQDTPRGVGLGVLPGTSVMQVAT 240
 QY 241 DEDDAIYTVNGVAVYSIHSQEPKDPHDMFTIHRSTGTISVISGGLDREKVPYTTTQA 300
 Db 241 DEDDAIYTVNGVAVYSIHSQEPKDPHDMFTIHRSTGTISVISGGLDREKVPYTTTQA 300
 QY 301 TMDGDSGTTTAVAVEILDANDNAPMFDPOKYEAHVPENAVGHEVQRLTVDLDAPNSP 360
 Db 301 TMDGDSGTTTAVAVEILDANDNAPMFDPOKYEAHVPENAVGHEVQRLTVDLDAPNSP 360
 QY 361 AWRATYILMGDDGDGHFTITTHPESNOGILTTKGLDPEAKNQHTLYVEVNEAPFVLKL 420
 Db 361 AWRATYILMGDDGDGHFTITTHPESNOGILTTKGLDPEAKNQHTLYVEVNEAPFVLKL 420
 QY 421 PTSTATIVVHVEDVNEAPVFPVPPSKVVEQEGIPTGEPVVCVVTABDPKQKISYRILR 480
 Db 421 PTSTATIVVHVEDVNEAPVFPVPPSKVVEQEGIPTGEPVVCVVTABDPKQKISYRILR 480
 QY 481 DPAGWLAMPDPSGQVAVGTGLDREDEQFVRNNIYEVWVLMNDGSPPTTGTGTLTLLID 540
 Db 481 DPAGWLAMPDPSGQVAVGTGLDREDEQFVRNNIYEVWVLMNDGSPPTTGTGTLTLLID 540
 QY 541 VNDHGVPPEPRQITICNQSPVRHVNITDKDLSPTSPFQALTDSDIYVTAENVNEGD 600
 Db 541 VNDHGVPPEPRQITICNQSPVRHVNITDKDLSPTSPFQALTDSDIYVTAENVNEGD 600
 QY 601 TVVLSLKXFLKQDITYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGQFIPVL 660
 Db 601 TVVLSLKXFLKQDITYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGQFIPVL 660
 QY 661 GAVLALLFLLVLLVLLVLRKKRKKIKFLLPDDTRDNVFFYGGEGGEDQYDITQLHR 720
 Db 661 GAVLALLFLLVLLVLLVLRKKRKKIKFLLPDDTRDNVFFYGGEGGEDQYDITQLHR 720

QY 721 GLEAPVLRNDVAPTIPTMYRPRPANPDIGNFIENLKAANTDPTAPPYDILLVF 780
 DB 721 GLEAPVLRNDVAPTIPTMYRPRPANPDIGNFIENLKAANTDPTAPPYDILLVF 780
 QY 781 DYEGSGDAASLSLSSASDQDDYDLNMGSRFKKLADMYGGGDD 829
 DB 781 DYEGSGDAASLSLSSASDQDDYDLNMGSRFKKLADMYGGGDD 829

RESULT 11

ABU56434
 ID ABU56434 standard; protein; 829 AA.
 XX AC
 XX AC ABU56434;
 XX DT 02-APR-2003 (first entry)
 XX DE Lung cancer-associated polypeptide #27.
 XX KW Lung cancer-associated polypeptide; cytostatic; emphysema;
 KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
 KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
 KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
 KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
 XX OS Unidentified.
 XX PN WO200286443-A2.
 XX PD 31-OCT-2002.
 XX XX
 XX 18-APR-2002; 2002WO-US012476.
 XX PR 18-APR-2001; 2001US-0284770P.
 PR 10-MAY-2001; 2001US-0290492P.
 PR 09-NOV-2001; 2001US-0339245P.
 PR 13-NOV-2001; 2001US-0350666P.
 PR 29-NOV-2001; 2001US-0334370P.
 PR 12-APR-2002; 2002US-0372246P.
 XX (EOSB-) EOS BIOTECHNOLOGY INC.
 XX PA Aziz N, Murray R;
 XX PI WPI; 2003-093161/08.
 XX DR N-PSDB; ABX76155.
 XX XX
 PT Detecting a lung cancer-associated transcript in a cell from a patient
 PT for treating lung cancer, by contacting a biological sample from the
 PT patient with a polynucleotide that exhibits increased or decreased
 PT expression in lung cancer.
 XX
 XX Claim 27; Page 210; 453pp; English.

XX The invention relates to a method for detecting a lung cancer-associated
 CC transcript in a cell from a patient, comprising contacting a biological
 CC sample from the patient with a polynucleotide that selectively hybridises
 CC to a sequence that is at least 80 % identical to a gene that exhibits
 CC increased or decreased expression in lung cancer samples. Lung cancer-
 CC associated polynucleotides and polypeptides are used for identifying a
 CC compound that modulates a lung cancer-associated polypeptide, for
 CC inhibiting proliferation of a lung cancer-associated cell to treat lung
 CC cancer in a patient and for treating a mammal having lung cancer by
 CC administering a modulatory compound identified. The methods are useful
 CC for treating lung cancer, such as small cell lung cancer, non-small cell
 CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,
 CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,
 CC hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and
 CC bronchiectasis. The genes, polynucleotides and polypeptides are useful
 CC for diagnostic purposes and as targets for screening for therapeutic
 CC compounds that modulate lung cancer, such as antibodies. Sequences
 CC ABU56408-ABU56745 represent lung cancer-associated polypeptides of the
 CC invention

XX Sequence 829 AA;
 SQ Query Match 87.8%; Score 728; DB 6; Length 829;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MGLPRGFLASLLLOVCWLOCAASEPCRAVFAEAETLEAGGAEQPGQALGVFVGCPCG 60
 DB 1 MGLPRGFLASLLLOVCWLOCAASEPCRAVFAEAETLEAGGAEQPGQALGVFVGCPCG 60
 QY 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKILRRHKRDVWVAPISVPENG 120
 DB 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPPSKILRRHKRDVWVAPISVPENG 120
 QY 121 KGPPQRLNOLKSNKORDTKIFYISITGPGADSPPEGVFAVEKETGMLLNKPLDREELAK 180
 DB 121 KGPPQRLNOLKSNKORDTKIFYISITGPGADSPPEGVFAVEKETGMLLNKPLDREELAK 180
 QY 181 YELFGHAVSNGASVEDPMNISIIITDQNDHKPKFTQDTFRGSLVLEGVLPSTVMQVAT 240
 DB 181 YELFGHAVSNGASVEDPMNISIIITDQNDHKPKFTQDTFRGSLVLEGVLPSTVMQVAT 240
 QY 241 DEDDAITYNGVWAYSIIHSQEPKDPHDLMTIHRSTGTISVISSGLDREKVPYTTIQA 300
 DB 241 DEDDAITYNGVWAYSIIHSQEPKDPHDLMTIHRSTGTISVISSGLDREKVPYTTIQA 300
 QY 301 TDMGDGSTTTAVAVVEILDANDNAPMFDQKYEAHVPENAVGHEVQRLTVDLDAPNSP 360
 DB 301 TDMGDGSTTTAVAVVEILDANDNAPMFDQKYEAHVPENAVGHEVQRLTVDLDAPNSP 360
 QY 361 AWRTYILMGDDGDHPTIITHPESNOGILTTTRKGLDFAKNOHTLYVEVTNAPFVKL 420
 DB 361 AWRTYILMGDDGDHPTIITHPESNOGILTTTRKGLDFAKNOHTLYVEVTNAPFVKL 420
 QY 421 PTSTATIIVHVEDVNEAPVFPVPSKVVEQEGPTGEPVCVYTAEDPKDKNQKISVRIILR 480
 DB 421 PTSTATIIVHVEDVNEAPVFPVPSKVVEQEGPTGEPVCVYTAEDPKDKNQKISVRIILR 480
 QY 481 DPAGWLAMDESDGQVTAAGTLDREDEQFVRNNTYEVNVLAMDNQSPPTTGTGTLTLTLD 540
 DB 481 DPAGWLAMDESDGQVTAAGTLDREDEQFVRNNTYEVNVLAMDNQSPPTTGTGTLTLTLD 540
 QY 541 VNDHGPVPEPRQITICNOSPVRHLNITDKLSPTSPFQAQLTDDSDIYWTAEVNEEGD 600
 DB 541 VNDHGPVPEPRQITICNOSPVRHLNITDKLSPTSPFQAQLTDDSDIYWTAEVNEEGD 600
 QY 601 TVVLSLKKFLKQDITYVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPWKGGFILPVL 660
 DB 601 TVVLSLKKFLKQDITYVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPWKGGFILPVL 660
 QY 661 GAVLALLFLLVLLVLRKKRKEPILLPDEDDTRDNVYFYEEGGGEEDQDITQLHR 720
 DB 661 GAVLALLFLLVLLVLRKKRKEPILLPDEDDTRDNVYFYEEGGGEEDQDITQLHR 720
 QY 721 GLEAPVLRNDVAPTIPTMYRPRPANPDIGNFIENLKAANTDPTAPPYDILLVF 780
 DB 721 GLEAPVLRNDVAPTIPTMYRPRPANPDIGNFIENLKAANTDPTAPPYDILLVF 780
 QY 781 DYEGSGDAASLSLSSASDQDDYDLNMGSRFKKLADMYGGGDD 829
 DB 781 DYEGSGDAASLSLSSASDQDDYDLNMGSRFKKLADMYGGGDD 829

RESULT 12

ABU56670
 ID ABU56670 standard; protein; 829 AA.
 XX AC
 XX AC ABU56670;
 XX DT 02-APR-2003 (first entry)
 XX DE Lung cancer-associated polypeptide #263.

XX Lung cancer-associated polypeptide; cytostatic; emphysema;
 KW antiinflammatory; antiasthmatic; non-small cell lung cancer; atelectasis;
 KW small cell lung cancer; benign lesion; precancerous lesion; bronchitis;
 KW chronic obstructive pulmonary disease; hypersensitivity pneumonitis;
 KW interstitial pulmonary fibrosis; fibrosis; asthma; bronchiectasis.
 XX
 OS Unidentified.
 XX
 XX WO200286443-A2.
 FN
 PD 31-OCT-2002.
 XX
 XX 18-APR-2002; 2002WO-US012476.
 PF
 XX 18-APR-2001; 2001US-0284770P.
 XX
 PR 10-MAY-2001; 2001US-0290492P.
 PR 09-NOV-2001; 2001US-0339245P.
 PR 13-NOV-2001; 2001US-0350666P.
 PR 29-NOV-2001; 2001US-0334370P.
 PR 12-APR-2002; 2002US-0372246P.
 XX
 XX (EOSB-) EOS BIOTECHNOLOGY INC.
 PA
 XX
 XX Aziz N, Murray R;
 PI
 XX
 DR WPI: 2003-093161/08.
 DR N-PSDB; ABX76399.
 XX
 XX Detecting a lung cancer-associated transcript in a cell from a patient
 PT for treating lung cancer, by contacting a biological sample from the
 PT patient with a polynucleotide that exhibits increased or decreased
 PT expression in lung cancer.
 XX
 XX Claim 27; Page 392; 453pp; English.
 PS
 XX The invention relates to a method for detecting a lung cancer-associated
 XX transcript in a cell from a patient, comprising contacting a biological
 CC sample from the patient with a polynucleotide that selectively hybridises
 CC to a sequence that is at least 80 % identical to a gene that exhibits
 CC increased or decreased expression in lung cancer samples. Lung cancer-
 CC associated polynucleotides and polypeptides are used for identifying a
 CC compound that modulates a lung cancer-associated polypeptide, for
 CC inhibiting proliferation of a lung cancer-associated cell to treat lung
 CC cancer in a patient and for treating a mammal having lung cancer by
 CC administering a modulatory compound identified. The methods are useful
 CC for treating lung cancer, such as small cell lung cancer, non-small cell
 CC lung cancer or other benign or precancerous lesions, e.g. atelectasis,
 CC emphysema, bronchitis, chronic obstructive pulmonary disease, fibrosis,
 CC hypersensitivity pneumonitis, interstitial pulmonary fibrosis, asthma and
 CC bronchiectasis. The genes, polynucleotides and polypeptides are useful
 CC for diagnostic purposes and as targets for screening for therapeutic
 CC compounds that modulate lung cancer, such as antibodies. Sequences
 CC ABU56408-ABU56745 represent lung cancer-associated polypeptides of the
 CC invention
 XX
 SQ Sequence 829 AA;
 Query Match 81.2%; Score 673; DB 6; Length 829;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 773; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 56 MGCFGQEPALFSTNDNDFVRNGETVQERRSLKERNPLKIFPSKRILRHKRDVWVAPIS 115
 DB 56 MGCFGQEPALFSTNDNDFVRNGETVQERRSLKERNPLKIFPSKRILRHKRDVWVAPIS 115
 QY 116 VPENGKGPFPORLNQKSNKDRDTKIFYSITGPGADSPPEGFAVEKETGWLILNKPLDR 175
 DB 116 VPENGKGPFPORLNQKSNKDRDTKIFYSITGPGADSPPEGFAVEKETGWLILNKPLDR 175
 QY 176 EETAKYELFGHAYSENGASVEDPMNISIIVTDQNDHKPKFTQDTFRGSLVGLPQTSVM 235
 DB 176 EETAKYELFGHAYSENGASVEDPMNISIIVTDQNDHKPKFTQDTFRGSLVGLPQTSVM 235

QY 236 QVTATDEDDAIYNGVWVAYSIHQSOPKOPHDLMFTHSRSTGTISVISGLDREKVPEYT 235
 DB 236 QVTATDEDDAIYNGVWVAYSIHQSOPKOPHDLMFTHSRSTGTISVISGLDREKVPEYT 235
 QY 296 LTIQATMDGSGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVAGHEVQRLTVDLD 355
 DB 296 LTIQATMDGSGSTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVAGHEVQRLTVDLD 355
 QY 356 APNSPAWRATYILMGDDGDHFTITTHPSNQILTRKGLDFAKNQHTLYVEVINEAP 415
 DB 356 APNSPAWRATYILMGDDGDHFTITTHPSNQILTRKGLDFAKNQHTLYVEVINEAP 415
 QY 416 FVLKLPSTSTATIVVHVVEDVNEAPVFPVPPSKVVEVQEGIPTEGPEVCVYTAEDDPKENQKIS 475
 DB 416 FVLKLPSTSTATIVVHVVEDVNEAPVFPVPPSKVVEVQEGIPTEGPEVCVYTAEDDPKENQKIS 475
 QY 476 YRILRDPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGLL 535
 DB 476 YRILRDPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGLL 535
 QY 536 LTLIDVNDHGPVPEPROITICNOSPVRHVNLNITDKLSPHTSPFOAQLTDDSDIYWTAEV 595
 DB 536 LTLIDVNDHGPVPEPROITICNOSPVRHVNLNITDKLSPHTSPFOAQLTDDSDIYWTAEV 595
 QY 596 NEEGDTVVLKXKFLKQDITYVHLSLSDHGNKQQLTVIRATVCDCHGHVETCPGPKMGKF 655
 DB 596 NEEGDTVVLKXKFLKQDITYVHLSLSDHGNKQQLTVIRATVCDCHGHVETCPGPKMGKF 655
 QY 656 ILPVLGAVTALLFLLVLLVLLVKKKIKKPKLLPEDDTEDNVYVYGEEGGEDDQDYDI 715
 DB 656 ILPVLGAVTALLFLLVLLVLLVKKKIKKPKLLPEDDTEDNVYVYGEEGGEDDQDYDI 715
 QY 716 TQHRGLEARPEVVLNRDVAFTIIPMYRPRPANPDEIGNFIENLKAANTDPTAPPYD 775
 DB 716 TQHRGLEARPEVVLNRDVAFTIIPMYRPRPANPDEIGNFIENLKAANTDPTAPPYD 775
 QY 776 TLLVDFYEGSGSDAASLSLSTSSASDQDQDYDLNENGSFVKLADMYGGGDD 829
 DB 776 TLLVDFYEGSGSDAASLSLSTSSASDQDQDYDLNENGSFVKLADMYGGGDD 829
 RESULT 13
 AAAY09093
 ID AAAY09093 standard; protein; 108 AA.
 XX
 AC AAAY09093;
 XX
 DT 07-JUL-1999 (first entry)
 XX
 DE Human P-cadherin extracellular domain 1.
 XX
 KW Cadherin; modulating agent; cadherin-mediated cell adhesion; cancer;
 KW drug delivery; bladder tumour; ovarian tumour; melanoma; cell adhesion;
 KW wound healing; neurite outgrowth; demyelinating neurological disease;
 KW multiple sclerosis; carcinoma; leukemia; melanoma; angiogenesis;
 KW apoptosis; diabetes; rheumatoid arthritis; immune system; pregnancy;
 KW vasopermeability; spinal cord injury; synaptic stability.
 XX
 OS Homo sapiens.
 XX
 PN WO9916791-A2.
 XX
 PD 08-APR-1999.
 XX
 PF 29-SEP-1998; 99WO-CA000902.
 XX
 PR 29-SEP-1997; 97US-00939853.
 XX
 PA (ADHE-) ADHEREX INC.
 XX
 PI Blaschuck OW, Gour BJ;
 XX

DR WPI; 1999-263686/22.
 XX Modulating cadherin-mediated cell adhesion useful for treating
 PT neurological disease and cancer.
 XX
 PS Disclosure; Fig 2; 148pp; English.
 XX
 CC The invention relates to methods using cadherin modulating agents,
 CC particularly peptides containing the sequence HAV or antibodies, for
 CC enhancing or inhibiting cadherin-mediated cell adhesion. The methods can
 CC be used to treat diseases or other conditions characterized by
 CC undesirable cell adhesion or to facilitate drug delivery to a specific
 CC tissue or tumour (such as bladder tumours, ovarian tumours or melanomas).
 CC They can also be used to enhance cell adhesion (e.g. supplement or
 CC replace stitches or to facilitate wound healing). They can also be used
 CC for enhancing and/or directing neurite outgrowth. The methods can also be
 CC used for treating demyelinating neurological disease, e.g. multiple
 CC sclerosis. The methods can also be used for e.g. enhancing drug delivery,
 CC treating cancers (such as carcinomas, leukemia or melanomas), inhibiting
 CC angiogenesis, enhancing adhesion of foreign tissue implanted in a mammal,
 CC inducing apoptosis in cadherin-expressing cells, e.g. for treating
 CC diabetes or rheumatoid arthritis, modulating the immune system, for
 CC preventing pregnancy, increasing vasopermeability, treating spinal cord
 CC injuries or inhibiting synaptic stability
 XX
 SQ Sequence 108 AA;

Query Match 13.0%; Score 108; DB 2; Length 108;
 Best Local Similarity 100.0%; Pred. No. 3.4e-93;
 Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 108 DWVAPISVPENGKGFPPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWL 167
 DB 1 DWVAPISVPENGKGFPPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWL 60
 QY 168 LLNKPLDREIAKYELFGHVSNGASVEDPMNISIIVTDQNDHKPKF 215
 DB 61 LLNKPLDREIAKYELFGHVSNGASVEDPMNISIIVTDQNDHKPKF 108

RESULT 14
 AAY17104
 ID AAY17104 standard; protein; 108 AA.
 AC AAY17104;
 DT 20-SEP-1999 (first entry)
 DE Human P-cadherin EC1 domain sequence.
 XX
 KW Cell adhesion modulation; CAM; synaptic stability; cadherin;
 KW cadherin-mediated adhesion; drug delivery; cell adhesion; tumour;
 KW wound healing; neurite outgrowth.
 XX
 OS Homo sapiens.
 XX
 FN WO9933875-A1.
 XX
 PD 08-JUL-1999.
 XX
 PF 23-DEC-1998; 98WO-CA001207.
 XX
 PR 23-DEC-1997; 97US-00996679.
 XX
 XX (UYMC-) UNIV MCGILL.
 PA Blaschuk OW, Gour BJ;
 XX WPI; 1999-430231/36.
 DR
 XX Cyclic peptide cell adhesion modulating agents, useful for modulating
 PT synaptic stability.
 XX

PS Disclosure; Fig 2; 144pp; English.
 XX
 CC The invention provides cyclic peptide cell adhesion modulating (CAM)
 CC agents that comprises a His-Ala-Val recognition sequence. Also provided
 CC is a method for inhibiting synaptic stability in a mammal that comprises
 CC administering to a mammal a therapeutically effective amount of a CAM
 CC agent that inhibits cadherin-mediated adhesion, where the agent comprises
 CC a cyclic peptide having a peptide ring, and where the sequence His-Ala-
 CC Val is present within the peptide ring. The cyclic peptides are cell
 CC adhesion modulating agents that inhibit cadherin-mediated adhesion. They
 CC can be used in a method for inhibiting synaptic stability in mammals. The
 CC agents can be used to treat diseases or other conditions characterized by
 CC undesirable cell adhesion or to facilitate drug delivery to a specific
 CC tissue or tumour. Alternatively the agents may be used to enhance cell
 CC adhesion (e.g. to supplement or replace stitches or to facilitate wound
 CC healing) or to enhance or direct neurite outgrowth
 XX
 SQ Sequence 108 AA;

Query Match 13.0%; Score 108; DB 2; Length 108;
 Best Local Similarity 100.0%; Pred. No. 3.4e-93;
 Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 108 DWVAPISVPENGKGFPPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWL 167
 DB 1 DWVAPISVPENGKGFPPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWL 60
 QY 168 LLNKPLDREIAKYELFGHVSNGASVEDPMNISIIVTDQNDHKPKF 215
 DB 61 LLNKPLDREIAKYELFGHVSNGASVEDPMNISIIVTDQNDHKPKF 108

RESULT 15
 AAY78184
 ID AAY78184 standard; protein; 108 AA.
 AC AAY78184;
 XX
 DT 03-MAY-2000 (first entry)
 XX
 DE Human P-cadherin EC1 domain SEQ ID NO:37.
 XX
 KW Cadherin; modulation; cell adhesion recognition; CAR; HAV-BF;
 KW HAV binding motif; antiapoptotic; cytostatic; cell adhesion;
 KW neurite outgrowth; Schwann cell migration; tumour; cancer; metastasis;
 KW apoptosis; angiogenesis; immune system; pregnancy; vasopermeability;
 KW demyelinating neurological disease.
 XX
 OS Homo sapiens.
 XX
 FN WO200002917-A2.
 XX
 PD 20-JAN-2000.
 XX
 PF 12-JUL-1999; 99WO-CA000627.
 XX
 PR 10-JUL-1998; 98US-00113977.
 XX
 XX (ADHE-) ADHEREX TECHNOLOGIES INC.
 XX
 PI Doherty P, Blaschuk OW, Gour BJ;
 XX WPI; 2000-160899/14.
 DR
 XX Modulating agents for treating autoimmune diseases, cancer, spinal cord
 PT injuries, and for increasing vasopermeability, inhibiting synaptic
 PT stability and preventing pregnancy.
 XX
 XX Disclosure; Fig 2; 144pp; English.
 XX
 CC The present invention describes a cell adhesion modulating agent (I)
 CC capable of binding to the cadherin cell adhesion recognition (CAR)
 CC sequence HAV, where the agent doesn't comprise an antibody or antigen-

CC binding fragment of it. (I) is used for inhibiting or enhancing cadherin
CC mediated functions like cell adhesion, neurite outgrowth, Schwann cell
CC migration and synaptic stability in cells preferably epithelial,
CC endothelial, neural, tumour cells and lymphocytes expressing cadherin E
CC or N. Inhibition of cadherin mediated cell adhesion by (I) is used in
CC reducing unwanted cellular adhesion, enhancing drug delivery through
CC skin, drug delivery to a tumour, treating cancer and/or inhibiting
CC metastasis, inducing apoptosis, inhibiting angiogenesis, modulating
CC immune system, preventing pregnancy, increasing vasopermeability,
CC inhibiting synaptic stability in a mammal. Enhancement is used for
CC facilitating wound healing, enhancing adhesion of foreign tissue implant,
CC enhancing and/or directing neurite outgrowth and treating spinal cord
CC injuries. (I) is used for treating demyelinating neurological diseases
CC and for facilitating blood sampling in a mammal. (I) is used for
CC detecting cadherin expressing cells in a sample by detecting the level of
CC antibody cadherin complex. AAy78148 to AAy78242 represent sequences used
CC in the exemplification of the present invention
XX
SQ Sequence 108 AA;

Query Match 13.0%; Score 108; DB 3; Length 108;
Best Local Similarity 100.0%; Pred. No. 3.4e-93;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 108 DWVAPISVPENGKGPFFQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWL 167
DB 1 DWVAPISVPENGKGPFFQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWL 60
QY 168 LLNKLDRREIAKYELFGHANSNGASVEDPMNISIIIVTDQNDHKPKF 215
DB 61 LLNKLDRREIAKYELFGHANSNGASVEDPMNISIIIVTDQNDHKPKF 108

Search completed: September 21, 2004, 22:23:04
Job time : 136 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004. Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:20:55 : Search time 34 Seconds
(without alignments)
1258.763 Million cell updates/sec

Title: US-09-916-849A-1
Perfect score: 829
Sequence: 1 MGLPRGPLASLLLLQVCWLQ.....NEWGRFKKLADMYGGCEDD 829

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0
Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/aaa/5A_COMB.pap:*
2: /cgn2_6/ptodata/2/aaa/5B_COMB.pap:*
3: /cgn2_6/ptodata/2/aaa/6A_COMB.pap:*
4: /cgn2_6/ptodata/2/aaa/6B_COMB.pap:*
5: /cgn2_6/ptodata/2/aaa/3CTUS_COMB.pap:*
6: /cgn2_6/ptodata/2/aaa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	13.0	108	3	US-08-893-534A-4
2	108	13.0	108	3	US-08-996-679-4
3	108	13.0	108	3	US-08-939-853A-6
4	108	13.0	108	3	US-09-115-395-4
5	108	13.0	108	3	US-09-113-977C-37
6	108	13.0	108	4	US-09-507-102-4
7	108	13.0	108	4	US-09-250-059-4
8	108	13.0	108	4	US-09-248-074-4
9	108	13.0	108	4	US-09-357-717-4
10	108	13.0	108	4	US-09-458-870-4
11	108	13.0	108	4	US-09-351-048A-37
12	108	13.0	108	4	US-09-248-015-4
13	108	13.0	108	4	US-09-544-782-4
14	106	12.8	106	3	US-09-113-977C-46
15	106	12.8	106	4	US-09-351-048A-46
16	51	6.2	822	2	US-08-474-067-9
17	51	6.2	822	2	US-08-474-068A-9
18	51	6.2	822	2	US-08-472-481-8
19	43	5.2	43	4	US-09-057-363C-56
20	43	5.2	43	4	US-09-265-107-56
21	22	2.7	107	1	US-08-326-117B-3
22	22	2.7	107	3	US-08-982-129-3
23	22	2.7	108	3	US-08-893-534A-5
24	22	2.7	108	3	US-08-996-679-5
25	22	2.7	108	3	US-08-939-853A-7
26	22	2.7	108	3	US-09-115-395-5
27	22	2.7	108	3	US-09-113-977C-38

28	22	2.7	108	4	US-09-507-102-5	Sequence 5, Appli
29	22	2.7	108	4	US-09-250-059-5	Sequence 5, Appli
30	22	2.7	108	4	US-09-248-074-5	Sequence 5, Appli
31	22	2.7	108	4	US-09-357-717-5	Sequence 5, Appli
32	22	2.7	108	4	US-09-458-870-5	Sequence 38, Appli
33	22	2.7	108	4	US-09-351-048A-38	Sequence 5, Appli
34	22	2.7	108	4	US-09-248-015-5	Sequence 5, Appli
35	22	2.7	108	4	US-09-544-782-5	Sequence 5, Appli
36	22	2.7	109	4	US-09-178-176B-8	Sequence 8, Appli
37	22	2.7	109	4	US-09-457-864-8	Sequence 8, Appli
38	22	2.7	109	4	US-09-457-865A-8	Sequence 8, Appli
39	17	2.1	837	2	US-08-474-067-7	Sequence 7, Appli
40	17	2.1	837	2	US-08-474-068A-7	Sequence 7, Appli
41	17	2.1	837	2	US-08-472-481-6	Sequence 6, Appli
42	15	1.8	884	2	US-08-474-067-8	Sequence 8, Appli
43	15	1.8	884	2	US-08-474-068A-8	Sequence 8, Appli
44	15	1.8	884	2	US-08-472-481-7	Sequence 7, Appli
45	14	1.7	43	4	US-09-057-363C-57	Sequence 57, Appli

ALIGNMENTS

RESULT 1
US-08-893-534A-4
; Sequence 4, Application US/08893534A
; Patent No. 6031072
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING
; TITLE OF INVENTION: CELL ADHESION
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,534A
; FILING DATE: 11-JUL-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 100086.401
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-893-534A-4

Query Match 13.0%; Score 108; DB 3; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	108	DWVAPISVPENGKPPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWL	167
Db	1	DWVAPISVPENGKPPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWL	60
Qy	168	LINKPDREETAKYELFGHAUSNGASVEDPMNISIIIVTDNDHKPF	215

Db 61 LLNKPLDREIAKYELFGHVSNGASVEDPWNISIIIVTDQNDHKPKF 108
RESULT 2
US-08-996-679-4
; Sequence 4, Application US/08996679
; Patent No. 6169071
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING
; CELL ADHESION
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 23-DEC-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 100086.401C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-996-679-4
Query Match 13.0%; Score 108; DB 3; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 108 DWVAPISVPENGKGFPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENGKGFPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60
Qy 168 LLNKPLDREIAKYELFGHVSNGASVEDPWNISIIIVTDQNDHKPKF 215
Db 61 LLNKPLDREIAKYELFGHVSNGASVEDPWNISIIIVTDQNDHKPKF 108
RESULT 3
US-08-939-853A-6
; Sequence 6, Application US/08939853A
; Patent No. 6203788
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR REGULATING
; CELL ADHESION
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington

; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/939,853A
; FILING DATE: 29-SEP-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 32,391
; REFERENCE/DOCKET NUMBER: 100086.402
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-939-853A-6
Query Match 13.0%; Score 108; DB 3; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 108 DWVAPISVPENGKGFPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENGKGFPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60
Qy 168 LLNKPLDREIAKYELFGHVSNGASVEDPWNISIIIVTDQNDHKPKF 215
Db 61 LLNKPLDREIAKYELFGHVSNGASVEDPWNISIIIVTDQNDHKPKF 108
RESULT 4
US-09-115-395-4
; Sequence 4, Application US/09115395A
; Patent No. 6207639
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NEURITE OUTGROWTH
; FILE REFERENCE: 100086.401C3
; CURRENT APPLICATION NUMBER: US/09/115,395A
; CURRENT FILING DATE: 1998-07-14
; EARLIER APPLICATION NUMBER: 08/996,679
; EARLIER FILING DATE: 1997-12-23
; EARLIER APPLICATION NUMBER: 08/893,534
; EARLIER FILING DATE: 1997-07-11
; EARLIER APPLICATION NUMBER: 60/021,612
; EARLIER FILING DATE: 1996-07-12
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-115-395-4
Query Match 13.0%; Score 108; DB 3; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 108 DWVAPISVPENGKGFPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENGKGFPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60
Qy 168 LLNKPLDREIAKYELFGHVSNGASVEDPWNISIIIVTDQNDHKPKF 215

Db 61 LNKPLDREETIAKYELFGHVSNGASVEDPMNISIIVTDQNDHKPKF 108
|||||

RESULT 5

US-09-113-977C-37
; Sequence 37, Application US/09113977C
; Patent No. 6277824
; GENERAL INFORMATION:
; APPLICANT: Doherty, Patrick
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING ADHESION MOLECULE
; FILE REFERENCE: 100086.403
; CURRENT APPLICATION NUMBER: US/09/113,977C
; CURRENT FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-113-977C-37

Query Match 13.0%; Score 108; DB 3; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 108 DWVAPISVPENGKGPFFQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENGKGPFFQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60
QY 168 LNKPLDREETIAKYELFGHVSNGASVEDPMNISIIVTDQNDHKPKF 215
Db 61 LNKPLDREETIAKYELFGHVSNGASVEDPMNISIIVTDQNDHKPKF 108

RESULT 6

US-09-507-102-4
; Sequence 4, Application US/09507102
; Patent No. 6326352
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING CELL ADHESION
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESS: SEED IP LAW GROUP PLLC
; STREET: 6300 Bank of America Bldg., 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/507,102
; FILING DATE: 17-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/893,534
; FILING DATE: 11-JUL-1997
; APPLICATION NUMBER: US 60/021,612
; FILING DATE: 12-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Christiansen, William T.
; REGISTRATION NUMBER: 44,614
; REFERENCE/DOCKET NUMBER: 100086.401C10

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-507-102-4

Query Match 13.0%; Score 108; DB 4; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 108 DWVAPISVPENGKGPFFQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENGKGPFFQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60
QY 168 LNKPLDREETIAKYELFGHVSNGASVEDPMNISIIVTDQNDHKPKF 215
Db 61 LNKPLDREETIAKYELFGHVSNGASVEDPMNISIIVTDQNDHKPKF 108

RESULT 7

US-09-250-059-4
; Sequence 4, Application US/09250059
; Patent No. 6333307
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NEURITE OUTGROWTH
; FILE REFERENCE: 100086.401C6
; CURRENT APPLICATION NUMBER: US/09/250,059
; CURRENT FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-250-059-4

Query Match 13.0%; Score 108; DB 4; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 108 DWVAPISVPENGKGPFFQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENGKGPFFQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60
QY 168 LNKPLDREETIAKYELFGHVSNGASVEDPMNISIIVTDQNDHKPKF 215
Db 61 LNKPLDREETIAKYELFGHVSNGASVEDPMNISIIVTDQNDHKPKF 108

RESULT 8

US-09-248-074-4
; Sequence 4, Application US/09248074
; Patent No. 6346512
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING CELL ADHESION
; FILE REFERENCE: 100086.401C5
; CURRENT APPLICATION NUMBER: US/09/248,074
; CURRENT FILING DATE: 1999-02-10
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 108
; TYPE: PRT

```
; ORGANISM: Homo sapiens
US-09-248-074-4
Query Match      13.0%; Score 108; DB 4; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 108 DWVAPISVPENKGGPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENKGGPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60

Qy 168 LNKPLDREIAKYELFGHVSNGASVEDPMNISIIITDQNDHKPKF 215
Db 61 LNKPLDREIAKYELFGHVSNGASVEDPMNISIIITDQNDHKPKF 108

RESULT 11
US-09-351-048A-37
; Sequence 37, Application US/09351048A
; Patent No. 6472368
; GENERAL INFORMATION:
; APPLICANT: Doherty, Patrick
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING ADHESION MOLECULE
; TITLE OF INVENTION: FUNCTION
; FILE REFERENCE: 100086.403C1
; CURRENT APPLICATION NUMBER: US/09/351.048A
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Solid Phase
; OTHER INFORMATION: Synthesis
US-09-351-048A-37

Query Match      13.0%; Score 108; DB 4; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 108 DWVAPISVPENKGGPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENKGGPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60

Qy 168 LNKPLDREIAKYELFGHVSNGASVEDPMNISIIITDQNDHKPKF 215
Db 61 LNKPLDREIAKYELFGHVSNGASVEDPMNISIIITDQNDHKPKF 108

RESULT 12
US-09-248-015-4
; Sequence 4, Application US/09248015
; Patent No. 6562786
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING APOPTOSIS
; FILE REFERENCE: 100086.401C4
; CURRENT APPLICATION NUMBER: US/09/248.015
; CURRENT FILING DATE: 1999-02-10
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-248-015-4

Query Match      13.0%; Score 108; DB 4; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 108 DWVAPISVPENKGGPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENKGGPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60

; ORGANISM: Homo sapiens
US-09-458-870-4
Query Match      13.0%; Score 108; DB 4; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 108 DWVAPISVPENKGGPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 167
Db 1 DWVAPISVPENKGGPPQRLNQLKSNKDRDTKIFYISITGPGADSPPEGVFAVEKETGWL 60

Qy 168 LNKPLDREIAKYELFGHVSNGASVEDPMNISIIITDQNDHKPKF 215
Db 61 LNKPLDREIAKYELFGHVSNGASVEDPMNISIIITDQNDHKPKF 108

RESULT 10
US-09-458-870-4
; Sequence 4, Application US/09458870
; Patent No. 6465427
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; APPLICANT: Farooqui, Riaz
; APPLICANT: Ali, Ammar
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING CELL ADHESION
; FILE REFERENCE: 100086.401C8
; CURRENT APPLICATION NUMBER: US/09/458.870
; CURRENT FILING DATE: 1999-12-10
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-458-870-4

Query Match      13.0%; Score 108; DB 4; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


QY 168 LINKPLDREIEIAKYELFGHVSNGASVEDPMNISIIIVTDQNDHKPKF 215
|
Db 61 LINKPLDREIEIAKYELFGHVSNGASVEDPMNISIIIVTDQNDHKPKF 108
|

RESULT 13

US-09-544-782-4
; Sequence 4, Application US/09544782
; Patent No. 6610821
; GENERAL INFORMATION:
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; APPLICANT: Farookhi, Riaz
; APPLICANT: Ali, Ammar
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING CELL ADHESION
; FILE REFERENCE: 100086.401C9
; CURRENT APPLICATION NUMBER: US/09/544,782
; CURRENT FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-544-782-4

Query Match 13.0%; Score 108; DB 4; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.1e-96;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 DWVAPISVPKNGKGFPPQRLNQLKSNKDRDKIFYISITGPGADSPPEGVFAVEKETGWL 167
|
Db 1 DWVAPISVPKNGKGFPPQRLNQLKSNKDRDKIFYISITGPGADSPPEGVFAVEKETGWL 60
|
QY 168 LINKPLDREIEIAKYELFGHVSNGASVEDPMNISIIIVTDQNDHKPKF 215
|
Db 61 LINKPLDREIEIAKYELFGHVSNGASVEDPMNISIIIVTDQNDHKPKF 108
|

RESULT 14

US-09-113-977C-46
; Sequence 46, Application US/09113977C
; Patent No. 6277824
; GENERAL INFORMATION:
; APPLICANT: Doherty, Patrick
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING ADHESION MOLECULE
; FILE REFERENCE: 100086.403
; CURRENT APPLICATION NUMBER: US/09/113,977C
; CURRENT FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-113-977C-46

Query Match 12.8%; Score 106; DB 3; Length 106;
Best Local Similarity 100.0%; Pred. No. 9.5e-95;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 441 VPPSKVVEVEGIGTGPVCVYTAEDPKENQKISYRILRDPAGWLAMPDPSGQVTA VGT 500
|
Db 1 VPPSKVVEVEGIGTGPVCVYTAEDPKENQKISYRILRDPAGWLAMPDPSGQVTA VGT 60
|
QY 501 LDREDEQFVRNNIYEVNVLAMDN GSPPTTGTGTLTLLTLDVNDHGP 546
|
Db 61 LDREDEQFVRNNIYEVNVLAMDN GSPPTTGTGTLTLLTLDVNDHGP 106
|

RESULT 15

US-09-351-048A-46
; Sequence 46, Application US/09351048A
; Patent No. 6472368
; GENERAL INFORMATION:
; APPLICANT: Doherty, Patrick
; APPLICANT: Blaschuk, Orest W.
; APPLICANT: Gour, Barbara J.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING ADHESION MOLECULE
; FILE REFERENCE: 100086.403C1
; CURRENT APPLICATION NUMBER: US/09/351,048A
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Solid Phase
; OTHER INFORMATION: Synthesis
US-09-351-048A-46

Query Match 12.8%; Score 106; DB 4; Length 106;
Best Local Similarity 100.0%; Pred. No. 9.5e-95;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 441 VPPSKVVEVEGIGTGPVCVYTAEDPKENQKISYRILRDPAGWLAMPDPSGQVTA VGT 500
|
Db 1 VPPSKVVEVEGIGTGPVCVYTAEDPKENQKISYRILRDPAGWLAMPDPSGQVTA VGT 60
|
QY 501 LDREDEQFVRNNIYEVNVLAMDN GSPPTTGTGTLTLLTLDVNDHGP 546
|
Db 61 LDREDEQFVRNNIYEVNVLAMDN GSPPTTGTGTLTLLTLDVNDHGP 106
|

Search completed: September 21, 2004, 22:27:27
Job time : 35 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 21, 2004, 22:26:02 ; Search time 141 Seconds
(without alignments)
1888.081 Million cell updates/sec

Title: US-09-916-849A-1
Perfect score: 829
Sequence: 1 MGLPRGPLASLLQLQVCWLQ.....NEWGSRFKLADMYGGGDD 829

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1342398 seqs, 32113274 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1342398

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/FCI_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/FCIU5_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	829	100.0	829	9	US-09-905-983-50
2	829	100.0	829	10	US-09-916-849A-1
3	829	100.0	829	10	US-09-975-723A-1
4	829	100.0	829	12	US-10-058-270A-126
5	829	100.0	829	14	US-10-174-677-21
6	829	100.0	829	14	US-10-158-123-1
7	829	100.0	829	15	US-10-295-027-783
8	829	100.0	829	15	US-10-295-027-800
9	829	100.0	829	15	US-10-295-027-863
10	829	100.0	829	15	US-10-295-027-896
11	728	87.8	829	12	US-10-229-345-16
12	728	87.8	829	12	US-10-274-177-16
13	728	87.8	829	14	US-10-301-822-18
14	728	87.8	829	15	US-10-295-027-338
15	728	87.8	829	15	US-10-295-027-1246

16	728	87.8	829	15	US-10-087-080-14	Sequence 14, Appl
17	728	87.8	829	16	US-10-650-112-16	Sequence 16, Appl
18	108	13.0	108	9	US-09-769-145-4	Sequence 4, Appl
19	108	13.0	108	10	US-09-778-026-6	Sequence 6, Appl
20	108	13.0	108	12	US-10-412-701-4	Sequence 4, Appl
21	108	13.0	108	12	US-10-105-008-4	Sequence 4, Appl
22	108	13.0	108	12	US-10-359-546-4	Sequence 4, Appl
23	108	13.0	108	13	US-10-006-982-4	Sequence 4, Appl
24	108	13.0	108	14	US-10-058-821-4	Sequence 4, Appl
25	108	13.0	108	14	US-10-193-653-37	Sequence 37, Appl
26	108	13.0	108	15	US-10-425-557-4	Sequence 4, Appl
27	108	13.0	108	16	US-10-632-678-4	Sequence 4, Appl
28	108	13.0	108	16	US-10-648-854-6	Sequence 46, Appl
29	106	12.8	106	14	US-10-193-653-46	Sequence 56, Appl
30	43	5.2	43	15	US-10-369-226-56	Sequence 56, Appl
31	22	2.7	107	10	US-09-457-865-3	Sequence 3, Appl
32	22	2.7	108	9	US-09-769-145-5	Sequence 5, Appl
33	22	2.7	108	10	US-09-778-026-7	Sequence 7, Appl
34	22	2.7	108	12	US-10-412-701-5	Sequence 5, Appl
35	22	2.7	108	12	US-10-105-008-5	Sequence 5, Appl
36	22	2.7	108	12	US-10-359-546-5	Sequence 5, Appl
37	22	2.7	108	13	US-10-006-982-5	Sequence 5, Appl
38	22	2.7	108	14	US-10-058-821-5	Sequence 5, Appl
39	22	2.7	108	14	US-10-193-653-38	Sequence 38, Appl
40	22	2.7	108	15	US-10-425-557-5	Sequence 5, Appl
41	22	2.7	108	16	US-10-632-678-5	Sequence 5, Appl
42	22	2.7	108	16	US-10-648-854-7	Sequence 7, Appl
43	22	2.7	109	14	US-10-187-088-8	Sequence 8, Appl
44	19	2.3	19	10	US-09-916-849A-4	Sequence 4, Appl
45	17	2.1	17	10	US-09-916-849A-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-09-905-983-50
; Sequence 50, Application US/09905983
; Patent No. US20020045591A1
; GENERAL INFORMATION:
; APPLICANT: Geiger, Benjamin
; APPLICANT: Ben-Ze'ev, Avri
; APPLICANT: Sadot, Einat
; FILE REFERENCE: 01/22326
; TITLE OF INVENTION: METHODS AND THERAPEUTIC COMPOSITIONS FOR TREATING CANCER
; CURRENT APPLICATION NUMBER: US/09/905.983
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-983-50

Query Match 100.0%; Score 829; DB 9; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MGLPRGPLASLLQLQVCWLQCAASEPCRAVPRAEVTLEAGGABOEPCQALGKVFMC	60
Db	1	MGLPRGPLASLLQLQVCWLQCAASEPCRAVPRAEVTLEAGGABOEPCQALGKVFMC	60
Qy	61	QEPALFSTDNDFTVRNGETVQERRSLKERNPLKIFPSKRLRHKRDWVAPISVPENG	120
Db	61	QEPALFSTDNDFTVRNGETVQERRSLKERNPLKIFPSKRLRHKRDWVAPISVPENG	120
Qy	121	KGPPFQRLNQLSKNRDRTKIFYITGADSPPEGVFAVEKFTGWLILNKLPLBREEIAK	180
Db	121	KGPPFQRLNQLSKNRDRTKIFYITGADSPPEGVFAVEKFTGWLILNKLPLBREEIAK	180
Qy	181	YELFGEHASENGASVEDPMNLSIIVTDNDHKPKFTQDTFEGSVLEGLVPGTSMQVMTAT	240

```
Db 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTPRGSVLEGLVPGTSVMQVAT 240
Qy 241 DEDDAIYTYNGVAVSYHSQEPKDPHLMFTIHRSTGTISVISGDLREKVPYTLTIOA 300
Db 241 DEDDAIYTYNGVAVSYHSQEPKDPHLMFTIHRSTGTISVISGDLREKVPYTLTIOA 300
Qy 301 TDMGDGSTTTAVAVBELDANAPMFDPKQYEAHVPENAVGHVORLTVTDLDAENSP 360
Db 301 TDMGDGSTTTAVAVBELDANAPMFDPKQYEAHVPENAVGHVORLTVTDLDAENSP 360
Qy 361 AWRATYILMGDDGDGHFTITTHPESNQGILTRKGLDPEAKNQHTLYVEVTNEAPFVLK 420
Db 361 AWRATYILMGDDGDGHFTITTHPESNQGILTRKGLDPEAKNQHTLYVEVTNEAPFVLK 420
Qy 421 PTSTATIVVHVEDVNEAPFVPPSKVVEQEGIPTGEPVCVYTAEDPKENQKISYRILR 480
Db 421 PTSTATIVVHVEDVNEAPFVPPSKVVEQEGIPTGEPVCVYTAEDPKENQKISYRILR 480
Qy 481 DPAGLWAMDPSGQVAVGLTLDREDEQFVRNNIYEVNVLAMDNQSPPTTGTGLLLTLID 540
Db 481 DPAGLWAMDPSGQVAVGLTLDREDEQFVRNNIYEVNVLAMDNQSPPTTGTGLLLTLID 540
Qy 541 VNDHGPVPEPRQITICNQSPVRHLNITDKDLSPHTSPPFOAQLTDDSDIYWTAEVNEEGD 600
Db 541 VNDHGPVPEPRQITICNQSPVRHLNITDKDLSPHTSPPFOAQLTDDSDIYWTAEVNEEGD 600
Qy 601 TVVLSLKKFLKQDYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
Db 601 TVVLSLKKFLKQDYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
Qy 661 GAVLALLFLLVLLVLRKKRKEPILLPEDDTRDNVFFYBEGGGEEDQDYDITQLHR 720
Db 661 GAVLALLFLLVLLVLRKKRKEPILLPEDDTRDNVFFYBEGGGEEDQDYDITQLHR 720
Qy 721 GLEARPEVLRNDVAPTIIPTMYRPRPNPDDEIGNFIENLKAANTDTPAPPYDILLVF 780
Db 721 GLEARPEVLRNDVAPTIIPTMYRPRPNPDDEIGNFIENLKAANTDTPAPPYDILLVF 780
Qy 781 DYEGSGDAASLSLTSASDQDDYDYLNEWGSRPKKLADMYGGGEDD 829
Db 781 DYEGSGDAASLSLTSASDQDDYDYLNEWGSRPKKLADMYGGGEDD 829

RESULT 2
US-09-916-849A-1
; Sequence 1, Application US/09916849A
; Publication No. US20030086934A1
; GENERAL INFORMATION:
; APPLICANT: Bostein, et al.
; TITLE OF INVENTION: Basal Markers in Breast Cancer and Related Reagents
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 2002850-0024
; CURRENT APPLICATION NUMBER: US/09/916,849A
; CURRENT FILING DATE: 2001-07-26
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 1
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Sequence of
; OTHER INFORMATION: Cadherin 3
US-09-916-849A-1

Query Match 100.0%; Score 829; DB 10; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGLPRGLASLLQLVCVILQCAASEPCRAVFPREAVTLEAGAEQEPQALGVFMGCPG 60
Db 1 MGLPRGLASLLQLVCVILQCAASEPCRAVFPREAVTLEAGAEQEPQALGVFMGCPG 60
```

```
Qy 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDVWVAPISVPENG 120
Db 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDVWVAPISVPENG 120
Qy 121 KPFPQRLNQKSNKDRDTKIFYSITGPADSPPEGVFAVEKETGWLKLLNKLDPRESIAK 180
Db 121 KPFPQRLNQKSNKDRDTKIFYSITGPADSPPEGVFAVEKETGWLKLLNKLDPRESIAK 180
Qy 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTPRGSVLEGLVPGTSVMQVAT 240
Db 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKFTQDTPRGSVLEGLVPGTSVMQVAT 240
Qy 241 DEDDAIYTYNGVAVSYHSQEPKDPHLMFTIHRSTGTISVISGDLREKVPYTLTIOA 300
Db 241 DEDDAIYTYNGVAVSYHSQEPKDPHLMFTIHRSTGTISVISGDLREKVPYTLTIOA 300
Qy 301 TDMGDGSTTTAVAVBELDANAPMFDPKQYEAHVPENAVGHVORLTVTDLDAENSP 360
Db 301 TDMGDGSTTTAVAVBELDANAPMFDPKQYEAHVPENAVGHVORLTVTDLDAENSP 360
Qy 361 AWRATYILMGDDGDGHFTITTHPESNQGILTRKGLDPEAKNQHTLYVEVTNEAPFVLK 420
Db 361 AWRATYILMGDDGDGHFTITTHPESNQGILTRKGLDPEAKNQHTLYVEVTNEAPFVLK 420
Qy 421 PTSTATIVVHVEDVNEAPFVPPSKVVEQEGIPTGEPVCVYTAEDPKENQKISYRILR 480
Db 421 PTSTATIVVHVEDVNEAPFVPPSKVVEQEGIPTGEPVCVYTAEDPKENQKISYRILR 480
Qy 481 DPAGLWAMDPSGQVAVGLTLDREDEQFVRNNIYEVNVLAMDNQSPPTTGTGLLLTLID 540
Db 481 DPAGLWAMDPSGQVAVGLTLDREDEQFVRNNIYEVNVLAMDNQSPPTTGTGLLLTLID 540
Qy 541 VNDHGPVPEPRQITICNQSPVRHLNITDKDLSPHTSPPFOAQLTDDSDIYWTAEVNEEGD 600
Db 541 VNDHGPVPEPRQITICNQSPVRHLNITDKDLSPHTSPPFOAQLTDDSDIYWTAEVNEEGD 600
Qy 601 TVVLSLKKFLKQDYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
Db 601 TVVLSLKKFLKQDYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKWGGFLLPVL 660
Qy 661 GAVLALLFLLVLLVLRKKRKEPILLPEDDTRDNVFFYBEGGGEEDQDYDITQLHR 720
Db 661 GAVLALLFLLVLLVLRKKRKEPILLPEDDTRDNVFFYBEGGGEEDQDYDITQLHR 720
Qy 721 GLEARPEVLRNDVAPTIIPTMYRPRPNPDDEIGNFIENLKAANTDTPAPPYDILLVF 780
Db 721 GLEARPEVLRNDVAPTIIPTMYRPRPNPDDEIGNFIENLKAANTDTPAPPYDILLVF 780
Qy 781 DYEGSGDAASLSLTSASDQDDYDYLNEWGSRPKKLADMYGGGEDD 829
Db 781 DYEGSGDAASLSLTSASDQDDYDYLNEWGSRPKKLADMYGGGEDD 829

RESULT 3
US-09-975-723A-1
; Sequence 1, Application US/09975723A
; Publication No. US20030108529A1
; GENERAL INFORMATION:
; APPLICANT: Nackman, Gary
; APPLICANT: Fotsy, Ramsey
; TITLE OF INVENTION: Improvement of Endothelial Cell-Cell
; TITLE OF INVENTION: Cohesion
; FILE REFERENCE: 601-1-101N
; CURRENT APPLICATION NUMBER: US/09/975,723A
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 60/241,216
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/243,693
; PRIOR FILING DATE: 2000-10-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 829
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-975-723A-1

Query Match      100.0%; Score 829; DB 10; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASLLLLQVCWLCQAASEPCRAVFREAEVTLLEAGGAEQEPGQALGKVFMGCPG 60
   |||||
Db 1 MGLPRGPLASLLLLQVCWLCQAASEPCRAVFREAEVTLLEAGGAEQEPGQALGKVFMGCPG 60
   |||||

QY 61 QEPALFSTDNDDFTVRNGETVQERSLKERNPLKIPPSKRIILRRHKRDVWVAPISVPENG 120
   |||||
Db 61 QEPALFSTDNDDFTVRNGETVQERSLKERNPLKIPPSKRIILRRHKRDVWVAPISVPENG 120
   |||||

QY 121 KGPPFQRLNQLKSNKDRDTKIFYITGPGADSPPEGVFAVEKETGMLLNKPLDREBIAK 180
   |||||
Db 121 KGPPFQRLNQLKSNKDRDTKIFYITGPGADSPPEGVFAVEKETGMLLNKPLDREBIAK 180
   |||||

QY 181 YELFGHAVSENGASVEDPMNISIIITDQNDHKPKETQDTRFGSVLEGLVPGTSVMQVAT 240
   |||||
Db 181 YELFGHAVSENGASVEDPMNISIIITDQNDHKPKETQDTRFGSVLEGLVPGTSVMQVAT 240
   |||||

QY 241 DEDDAIYTYNGVAVSIHQEPKDPHDLMTFTHRSTGTISVSSGLDREKVEYTLTIQA 300
   |||||
Db 241 DEDDAIYTYNGVAVSIHQEPKDPHDLMTFTHRSTGTISVSSGLDREKVEYTLTIQA 300
   |||||

QY 301 TMDGDSGTTTAVAVVEILDANDNAPMFPDQKYEAHVPENAVGHEVQRLTVDLDAENSP 360
   |||||
Db 301 TMDGDSGTTTAVAVVEILDANDNAPMFPDQKYEAHVPENAVGHEVQRLTVDLDAENSP 360
   |||||

QY 361 AWRATYILMGDDGDGHFTITTHPESNOGILTKRGLDPEAKNOHTLYEVNTNEAPFVLKL 420
   |||||
Db 361 AWRATYILMGDDGDGHFTITTHPESNOGILTKRGLDPEAKNOHTLYEVNTNEAPFVLKL 420
   |||||

QY 421 PTSTATIVVHVEDVNEAPVFVPPSKVVEQSGIPTGEPVCVYTAEDPKENQKISYRIIL 480
   |||||
Db 421 PTSTATIVVHVEDVNEAPVFVPPSKVVEQSGIPTGEPVCVYTAEDPKENQKISYRIIL 480
   |||||

QY 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLMNDGSPPTGTGTLTLLID 540
   |||||
Db 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLMNDGSPPTGTGTLTLLID 540
   |||||

QY 541 VNDHGVPVPRQITICNQSPVRHLNITDKLSPTSPPFQAQLTDDSDIYWTAEVNEEGD 600
   |||||
Db 541 VNDHGVPVPRQITICNQSPVRHLNITDKLSPTSPPFQAQLTDDSDIYWTAEVNEEGD 600
   |||||

QY 601 TVVLSLKKFLKQDITYDVHLSLSDHGNKEQITVIRATVCDCHGVETCPGPKGGFILPVL 660
   |||||
Db 601 TVVLSLKKFLKQDITYDVHLSLSDHGNKEQITVIRATVCDCHGVETCPGPKGGFILPVL 660
   |||||

QY 661 GAVIALLFLLVLLLVRRKRKIKEPLLPEDDTRDNVFFYEGEGGEDDQYDITQLHR 720
   |||||
Db 661 GAVIALLFLLVLLLVRRKRKIKEPLLPEDDTRDNVFFYEGEGGEDDQYDITQLHR 720
   |||||

QY 721 GLEAREVVLNDVAPTIPTMYRPRRANPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
   |||||
Db 721 GLEAREVVLNDVAPTIPTMYRPRRANPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
   |||||

QY 781 DYEGSGDAASLSLTSASDQDQDYDLNWSGRFELKADMYGGGEDD 829
   |||||
Db 781 DYEGSGDAASLSLTSASDQDQDYDLNWSGRFELKADMYGGGEDD 829
   |||||
```

RESULT 4

```
US-10-058-270A-126
; Sequence 126, Application US/10058270A
; Publication No. US20040029114A1
; GENERAL INFORMATION:
; APPLICANT: Mack, David H.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Afar, Daniel
```

```
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of diagnosis of Breast Cancer, Compositions and
; FILE OF INVENTION: Methods of Screening for Modulators of Breast Cancer
; FILE REFERENCE: 018501-005210US
; CURRENT APPLICATION NUMBER: US/10/058,270A
; CURRENT FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: US 60/263,965
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: US 60/265,928
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: US 09/829,472
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: US 60/282,698
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: US 60/288,590
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/294,443
; PRIOR FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 126
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-058-270A-126
```

```
Query Match      100.0%; Score 829; DB 12; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASLLLLQVCWLCQAASEPCRAVFREAEVTLLEAGGAEQEPGQALGKVFMGCPG 60
   |||||
Db 1 MGLPRGPLASLLLLQVCWLCQAASEPCRAVFREAEVTLLEAGGAEQEPGQALGKVFMGCPG 60
   |||||

QY 61 QEPALFSTDNDDFTVRNGETVQERSLKERNPLKIPPSKRIILRRHKRDVWVAPISVPENG 120
   |||||
Db 61 QEPALFSTDNDDFTVRNGETVQERSLKERNPLKIPPSKRIILRRHKRDVWVAPISVPENG 120
   |||||

QY 121 KGPPFQRLNQLKSNKDRDTKIFYITGPGADSPPEGVFAVEKETGMLLNKPLDREBIAK 180
   |||||
Db 121 KGPPFQRLNQLKSNKDRDTKIFYITGPGADSPPEGVFAVEKETGMLLNKPLDREBIAK 180
   |||||

QY 181 YELFGHAVSENGASVEDPMNISIIITDQNDHKPKETQDTRFGSVLEGLVPGTSVMQVAT 240
   |||||
Db 181 YELFGHAVSENGASVEDPMNISIIITDQNDHKPKETQDTRFGSVLEGLVPGTSVMQVAT 240
   |||||

QY 241 DEDDAIYTYNGVAVSIHQEPKDPHDLMTFTHRSTGTISVSSGLDREKVEYTLTIQA 300
   |||||
Db 241 DEDDAIYTYNGVAVSIHQEPKDPHDLMTFTHRSTGTISVSSGLDREKVEYTLTIQA 300
   |||||

QY 301 TMDGDSGTTTAVAVVEILDANDNAPMFPDQKYEAHVPENAVGHEVQRLTVDLDAENSP 360
   |||||
Db 301 TMDGDSGTTTAVAVVEILDANDNAPMFPDQKYEAHVPENAVGHEVQRLTVDLDAENSP 360
   |||||

QY 361 AWRATYILMGDDGDGHFTITTHPESNOGILTKRGLDPEAKNOHTLYEVNTNEAPFVLKL 420
   |||||
Db 361 AWRATYILMGDDGDGHFTITTHPESNOGILTKRGLDPEAKNOHTLYEVNTNEAPFVLKL 420
   |||||

QY 421 PTSTATIVVHVEDVNEAPVFVPPSKVVEQSGIPTGEPVCVYTAEDPKENQKISYRIIL 480
   |||||
Db 421 PTSTATIVVHVEDVNEAPVFVPPSKVVEQSGIPTGEPVCVYTAEDPKENQKISYRIIL 480
   |||||

QY 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLMNDGSPPTGTGTLTLLID 540
   |||||
Db 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLMNDGSPPTGTGTLTLLID 540
   |||||

QY 541 VNDHGVPVPRQITICNQSPVRHLNITDKLSPTSPPFQAQLTDDSDIYWTAEVNEEGD 600
   |||||
Db 541 VNDHGVPVPRQITICNQSPVRHLNITDKLSPTSPPFQAQLTDDSDIYWTAEVNEEGD 600
   |||||

QY 601 TVVLSLKKFLKQDITYDVHLSLSDHGNKEQITVIRATVCDCHGVETCPGPKGGFILPVL 660
   |||||
Db 601 TVVLSLKKFLKQDITYDVHLSLSDHGNKEQITVIRATVCDCHGVETCPGPKGGFILPVL 660
   |||||
```

```
QY 661 GAVLALLFLLLVLLVLLVLRKKRKKIPELPLLPEDDTRDNVFFYEGGGEEDQDYDITQLHR 720
Db 661 GAVLALLFLLLVLLVLLVLRKKRKKIPELPLLPEDDTRDNVFFYEGGGEEDQDYDITQLHR 720
QY 721 GLEARPEVLNVDVAPTIIPMPYRPRPNDPDEIGNFIENKAANTDTPAPPYDITLLVF 780
Db 721 GLEARPEVLNVDVAPTIIPMPYRPRPNDPDEIGNFIENKAANTDTPAPPYDITLLVF 780
QY 781 DYEGSGDAASLSLTSASDQDDYDLNWSRPFKKLADMYGGGEDD 829
Db 781 DYEGSGDAASLSLTSASDQDDYDLNWSRPFKKLADMYGGGEDD 829

RESULT 5
US-10-174-677-21
; Sequence 21, Application US/10174677
; Publication No. US20030190704A1
; GENERAL INFORMATION:
; APPLICANT: Xie, Ting
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ANCHORING STEM CELLS IN A MICROENVIR
; FILE REFERENCE: 40716(IP-012)
; CURRENT APPLICATION NUMBER: US/10/174,677
; CURRENT FILING DATE: 2002-06-19
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 21
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-174-677-21

Query Match 100.0%; Score 829; DB 14; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASLLLLQVCWLQCAASEPCRAVFREAEVTLLEAGAEQBPQALGVFMGCPG 60
Db 1 MGLPRGPLASLLLLQVCWLQCAASEPCRAVFREAEVTLLEAGAEQBPQALGVFMGCPG 60
QY 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPFSKRILRRHRKRDWVAPISVPENG 120
Db 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPFSKRILRRHRKRDWVAPISVPENG 120
QY 121 KGFPORLNQKSKRDRTKIFYSITGPGADSPPEGVFAVEKETGWLILNKLPLDREEIAK 180
Db 121 KGFPORLNQKSKRDRTKIFYSITGPGADSPPEGVFAVEKETGWLILNKLPLDREEIAK 180
QY 181 YELFGHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTRFGSVLEGLVPGTSVMQVAT 240
Db 181 YELFGHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTRFGSVLEGLVPGTSVMQVAT 240
QY 241 DEDDAIYTVNGVAVSIHSGEPKDPHDLMTIHRSTGTISVSSGLDREKVEPYTLTIOA 300
Db 241 DEDDAIYTVNGVAVSIHSGEPKDPHDLMTIHRSTGTISVSSGLDREKVEPYTLTIOA 300
QY 301 TMDGSGSTTTAVAVVEILDANDNAPMFPDQKYEAHVAVPENAVGHEVQRLTVDLDAPNSP 360
Db 301 TMDGSGSTTTAVAVVEILDANDNAPMFPDQKYEAHVAVPENAVGHEVQRLTVDLDAPNSP 360
QY 361 AWRATYILMGDDGDHFTTTHPESNOGILTRKGLDPEAKNQHTLYEVETNEAPFVLKL 420
Db 361 AWRATYILMGDDGDHFTTTHPESNOGILTRKGLDPEAKNQHTLYEVETNEAPFVLKL 420
QY 421 PTSTAIIVHVEDVNEAPVFPVPSKVVEQEGIPTEGVCVYTAEDDPKENGKISYRILR 480
Db 421 PTSTAIIVHVEDVNEAPVFPVPSKVVEQEGIPTEGVCVYTAEDDPKENGKISYRILR 480
QY 481 DPAGWLAMPDSQVAVGTLDREDEQFVRNNIYEVMLVAMNGSPPTGTGTLTLTID 540
Db 481 DPAGWLAMPDSQVAVGTLDREDEQFVRNNIYEVMLVAMNGSPPTGTGTLTLTID 540
QY 541 VNDHGVPPEPRQITICNQSPVRHVLNITDKDLSPHTSPPQAQLTDDSDIYWTAEVNEGSD 600
```

```
Db 541 VNDHGVPPEPRQITICNQSPVRHVLNITDKDLSPHTSPPQAQLTDDSDIYWTAEVNEGSD 600
QY 601 TVVLSLKKFLKQDYYDHLVLSLSDHGKKEQLTVIRATVCDCHGHVETCPGPKGGLPVL 660
Db 601 TVVLSLKKFLKQDYYDHLVLSLSDHGKKEQLTVIRATVCDCHGHVETCPGPKGGLPVL 660
QY 661 GAVLALLFLLLVLLVLLVLRKKRKKIPELPLLPEDDTRDNVFFYEGGGEEDQDYDITQLHR 720
Db 661 GAVLALLFLLLVLLVLLVLRKKRKKIPELPLLPEDDTRDNVFFYEGGGEEDQDYDITQLHR 720
QY 721 GLEARPEVLNVDVAPTIIPMPYRPRPNDPDEIGNFIENKAANTDTPAPPYDITLLVF 780
Db 721 GLEARPEVLNVDVAPTIIPMPYRPRPNDPDEIGNFIENKAANTDTPAPPYDITLLVF 780
QY 781 DYEGSGDAASLSLTSASDQDDYDLNWSRPFKKLADMYGGGEDD 829
Db 781 DYEGSGDAASLSLTSASDQDDYDLNWSRPFKKLADMYGGGEDD 829

RESULT 6
US-10-158-123-1
; Sequence 1, Application US/10158123
; Publication No. US20030194406A1
; GENERAL INFORMATION:
; APPLICANT: Reinhard, Christoph
; APPLICANT: Klinger, Julie
; APPLICANT: Jefferson, Ann
; APPLICANT: Escobedo, Jaime
; APPLICANT: Randazzo, Fillipo
; APPLICANT: Winter, Jill
; APPLICANT: Goodson, Robert
; APPLICANT: Qi, Weimin
; TITLE OF INVENTION: P-Cadherin as a Target for Anti-Cancer
; FILE REFERENCE: 35784/258994
; CURRENT APPLICATION NUMBER: US/10/158,123
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: 60/294,225
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-158-123-1

Query Match 100.0%; Score 829; DB 14; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGPLASLLLLQVCWLQCAASEPCRAVFREAEVTLLEAGAEQBPQALGVFMGCPG 60
Db 1 MGLPRGPLASLLLLQVCWLQCAASEPCRAVFREAEVTLLEAGAEQBPQALGVFMGCPG 60
QY 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPFSKRILRRHRKRDWVAPISVPENG 120
Db 61 QEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIPFSKRILRRHRKRDWVAPISVPENG 120
QY 121 KGFPORLNQKSKRDRTKIFYSITGPGADSPPEGVFAVEKETGWLILNKLPLDREEIAK 180
Db 121 KGFPORLNQKSKRDRTKIFYSITGPGADSPPEGVFAVEKETGWLILNKLPLDREEIAK 180
QY 181 YELFGHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTRFGSVLEGLVPGTSVMQVAT 240
Db 181 YELFGHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTRFGSVLEGLVPGTSVMQVAT 240
QY 241 DEDDAIYTVNGVAVSIHSGEPKDPHDLMTIHRSTGTISVSSGLDREKVEPYTLTIOA 300
Db 241 DEDDAIYTVNGVAVSIHSGEPKDPHDLMTIHRSTGTISVSSGLDREKVEPYTLTIOA 300
QY 301 TMDGSGSTTTAVAVVEILDANDNAPMFPDQKYEAHVAVPENAVGHEVQRLTVDLDAPNSP 360
```

```

Db 301 TMDGSGSTTTAVAVVILNDANDNAPMFDPOKYEAHVPENAVGHEVQLTVDLDAFNSP 360
Qy 361 AWRATYLMGGDGDGHFTITTHPESNOGILTRKGLDPEAKNQHTLYVEVNEAPFVLKL 420
Db 361 AWRATYLMGGDGDGHFTITTHPESNOGILTRKGLDPEAKNQHTLYVEVNEAPFVLKL 420
Qy 421 PTSTATIVVHVEDVNEAPVFPVPSKVVEVQEGIPTEGPEVCVVTAEADPKENKISYRILR 480
Db 421 PTSTATIVVHVEDVNEAPVFPVPSKVVEVQEGIPTEGPEVCVVTAEADPKENKISYRILR 480
Qy 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVVWLAMDNGSPPTTGTGLTLLTLD 540
Db 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVVWLAMDNGSPPTTGTGLTLLTLD 540
Qy 541 VNDHGPVPEPRQITTCNQSPVRHVNLITDKDLSPTSPFQAQLTDDSDIYWTAEVNEEGD 600
Db 541 VNDHGPVPEPRQITTCNQSPVRHVNLITDKDLSPTSPFQAQLTDDSDIYWTAEVNEEGD 600
Qy 601 TVVLSLKKFLKQDVTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGGFILLPVL 660
Db 601 TVVLSLKKFLKQDVTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGGFILLPVL 660
Qy 661 GAVLALLFLLVLLVLLVLRKKRIKEPILLPEDDTRDNVFFYGEEGGEDDQDYITQLHR 720
Db 661 GAVLALLFLLVLLVLLVLRKKRIKEPILLPEDDTRDNVFFYGEEGGEDDQDYITQLHR 720
Qy 721 GLEAREPVVLRNDVAPTIPTPMYRPRANPDEIGNFIENLKAANTDPTAPPDYTLVLF 780
Db 721 GLEAREPVVLRNDVAPTIPTPMYRPRANPDEIGNFIENLKAANTDPTAPPDYTLVLF 780
Qy 781 DYEGSGSDAASLSLTSASSDQDDQDYDLNENWGSRPFKKLADMYGGGEDD 829
Db 781 DYEGSGSDAASLSLTSASSDQDDQDYDLNENWGSRPFKKLADMYGGGEDD 829

RESULT 7
US-10-295-027-783
; Sequence 783, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Nacasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynnne, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250

```

```

; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 783
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-783

Query Match      100.0%; Score 829; DB 15; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGLPRGPLASLLLOVCWLOCAASEPCRAVFRAEAVTLEAGGAQOEFGOALGVFMGCPG 60
Db 1 MGLPRGPLASLLLOVCWLOCAASEPCRAVFRAEAVTLEAGGAQOEFGOALGVFMGCPG 60
Qy 61 QEPALFSTDNDNDFTVRNGETVQERRSLKERNPLKIPPSKILRRHKRDVVAIPSPENG 120
Db 61 QEPALFSTDNDNDFTVRNGETVQERRSLKERNPLKIPPSKILRRHKRDVVAIPSPENG 120
Qy 121 KGPPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLANKPLDREBIAK 180
Db 121 KGPPQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLANKPLDREBIAK 180
Qy 181 YELFGHAVSENGASVEDPMNISIIVTQNDHKPKFTQDTFRGSLVGLPCTSMQVAT 240
Db 181 YELFGHAVSENGASVEDPMNISIIVTQNDHKPKFTQDTFRGSLVGLPCTSMQVAT 240
Qy 241 DEDDAIYTVNGVAVYSIHSQEPKDPHLMFTIHRSTGTISVISGLDREKVPETLTIQA 300
Db 241 DEDDAIYTVNGVAVYSIHSQEPKDPHLMFTIHRSTGTISVISGLDREKVPETLTIQA 300
Qy 301 TMDGSGSTTTAVAVVILNDANDNAPMFDPOKYEAHVPENAVGHEVQLTVDLDAFNSP 360
Db 301 TMDGSGSTTTAVAVVILNDANDNAPMFDPOKYEAHVPENAVGHEVQLTVDLDAFNSP 360
Qy 361 AWRATYLMGGDGDGHFTITTHPESNOGILTRKGLDPEAKNQHTLYVEVNEAPFVLKL 420
Db 361 AWRATYLMGGDGDGHFTITTHPESNOGILTRKGLDPEAKNQHTLYVEVNEAPFVLKL 420
Qy 421 PTSTATIVVHVEDVNEAPVFPVPSKVVEVQEGIPTEGPEVCVVTAEADPKENKISYRILR 480
Db 421 PTSTATIVVHVEDVNEAPVFPVPSKVVEVQEGIPTEGPEVCVVTAEADPKENKISYRILR 480
Qy 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVVWLAMDNGSPPTTGTGLTLLTLD 540
Db 481 DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVVWLAMDNGSPPTTGTGLTLLTLD 540
Qy 541 VNDHGPVPEPRQITTCNQSPVRHVNLITDKDLSPTSPFQAQLTDDSDIYWTAEVNEEGD 600
Db 541 VNDHGPVPEPRQITTCNQSPVRHVNLITDKDLSPTSPFQAQLTDDSDIYWTAEVNEEGD 600
Qy 601 TVVLSLKKFLKQDVTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGGFILLPVL 660
Db 601 TVVLSLKKFLKQDVTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGGFILLPVL 660
Qy 661 GAVLALLFLLVLLVLLVLRKKRIKEPILLPEDDTRDNVFFYGEEGGEDDQDYITQLHR 720
Db 661 GAVLALLFLLVLLVLLVLRKKRIKEPILLPEDDTRDNVFFYGEEGGEDDQDYITQLHR 720
Qy 721 GLEAREPVVLRNDVAPTIPTPMYRPRANPDEIGNFIENLKAANTDPTAPPDYTLVLF 780
Db 721 GLEAREPVVLRNDVAPTIPTPMYRPRANPDEIGNFIENLKAANTDPTAPPDYTLVLF 780
Qy 781 DYEGSGSDAASLSLTSASSDQDDQDYDLNENWGSRPFKKLADMYGGGEDD 829
Db 781 DYEGSGSDAASLSLTSASSDQDDQDYDLNENWGSRPFKKLADMYGGGEDD 829

```

```
RESULT 8
US-10-295-027-800
; Sequence 800, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 800
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-800

Query Match
Best Local Similarity 100.0%; Score 829; DB 15; Length 829;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLPRGLASLLLLQVCWLQCAASEPCRAVFREAEVTLSEAGAEPPGALGKVFMGCPG 60
DB 1 MGLPRGLASLLLLQVCWLQCAASEPCRAVFREAEVTLSEAGAEPPGALGKVFMGCPG 60
QY 61 QEPALFSTNDFTVRNGETVQERRSLKERNPLKIPPSKRIILRRHKRWVAPISPENG 120
DB 61 QEPALFSTNDFTVRNGETVQERRSLKERNPLKIPPSKRIILRRHKRWVAPISPENG 120
QY 121 KGFPQRLNQLSKNRDRTKIFYSITPGADSPPEGVFAVEKETGWLNLKPLDREBIAK 180
DB 121 KGFPQRLNQLSKNRDRTKIFYSITPGADSPPEGVFAVEKETGWLNLKPLDREBIAK 180
QY 181 YELFGHAVSENGASVEDPNWISLIYTDQNDHKPKFTQDTFRGSVLEGLPGTSVMQVTAT 240
DB 181 YELFGHAVSENGASVEDPNWISLIYTDQNDHKPKFTQDTFRGSVLEGLPGTSVMQVTAT 240
QY 241 DEDDAIYTYNGVAVYIHSQEPKDPHDLNFTTHRSITGTSIVSSGLDREKVPYTLTQA 300
DB 241 DEDDAIYTYNGVAVYIHSQEPKDPHDLNFTTHRSITGTSIVSSGLDREKVPYTLTQA 300

QY 301 TDMGDGSGTTTAVAVVEILDANDANPMPFDPOKYEAHVPEANAVGHEVORLTVTDLDAENSP 360
DB 301 TDMGDGSGTTTAVAVVEILDANDANPMPFDPOKYEAHVPEANAVGHEVORLTVTDLDAENSP 360
QY 361 AWRATYILMGDDGDHFTITTHPESNQIGILTRKGLDPEAKNQHTLYVEVTNEAPFVVKL 420
DB 361 AWRATYILMGDDGDHFTITTHPESNQIGILTRKGLDPEAKNQHTLYVEVTNEAPFVVKL 420
QY 421 PTSTATIVVHVEDVNEAPVFPVPSKVVEQEGIPTGEPVCVYTAEDDPDKENOKISVILR 480
DB 421 PTSTATIVVHVEDVNEAPVFPVPSKVVEQEGIPTGEPVCVYTAEDDPDKENOKISVILR 480
QY 481 DPAGWLAMPDPSGGVTAAGTLDREDEQFVRNNIYEVNVLAMDNGSPPTTGTGTLTLLID 540
DB 481 DPAGWLAMPDPSGGVTAAGTLDREDEQFVRNNIYEVNVLAMDNGSPPTTGTGTLTLLID 540
QY 541 VNDHGPVPEPRQITICNQSPVRHVLTITDKLSHTSPFQAQLTDDSDIYWTAEVNEEGD 600
DB 541 VNDHGPVPEPRQITICNQSPVRHVLTITDKLSHTSPFQAQLTDDSDIYWTAEVNEEGD 600
QY 601 TVVLSLKKFLKQDTYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPWKGGFILPVL 660
DB 601 TVVLSLKKFLKQDTYDVHLSLSDHGNKEQLTVIRATVCDCHGVETCPGPWKGGFILPVL 660
QY 661 GAVLALLFLLVLLVLLVLRKKRKEPILLPEDDTRDNVFFYGGEGGEDQDYDITOLHR 720
DB 661 GAVLALLFLLVLLVLLVLRKKRKEPILLPEDDTRDNVFFYGGEGGEDQDYDITOLHR 720
QY 721 GLEARPEVLRNDVAPTIITPMYRPRPANDPDEIGNFIIENLKAANTDPTAPPYDTLLVF 780
DB 721 GLEARPEVLRNDVAPTIITPMYRPRPANDPDEIGNFIIENLKAANTDPTAPPYDTLLVF 780
QY 781 DYEGSGSDAASLSLSTSSASDQDQDYDLNENWGSRFKXKLADMYGGGEDD 829
DB 781 DYEGSGSDAASLSLSTSSASDQDQDYDLNENWGSRFKXKLADMYGGGEDD 829

RESULT 9
US-10-295-027-863
; Sequence 863, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
```


;; PRIOR APPLICATION NUMBER: US 60/355,250
;; PRIOR FILING DATE: 2002-02-08
;; PRIOR APPLICATION NUMBER: US 60/356,714
;; PRIOR FILING DATE: 2002-02-13
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 1386
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 863
;; LENGTH: 829
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-295-027-863

Query Match 100.0%; Score 829; DB 15; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MGLPRGLASLLLLQVCLWQCAASEPCRAVFRFAEAVTLFAGAGAEQPGALGKVFMC	60
Db	1	MGLPRGLASLLLLQVCLWQCAASEPCRAVFRFAEAVTLFAGAGAEQPGALGKVFMC	60
QY	61	QEPALFSTDNDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAPISVP	120
Db	61	QEPALFSTDNDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAPISVP	120
QY	121	KGFPFQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLILNKP	180
Db	121	KGFPFQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLILNKP	180
QY	181	YELFGHVSNGASVEDPMNIIIVTDQNDHKPKFTQDTRFGSVLEGLVPGTSMQV	240
Db	181	YELFGHVSNGASVEDPMNIIIVTDQNDHKPKFTQDTRFGSVLEGLVPGTSMQV	240
QY	241	DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTHRSTGTSIVSSGLDREKVP	300
Db	241	DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTHRSTGTSIVSSGLDREKVP	300
QY	301	TMMDGGSTTTAVAVVEILDANDNAPMDPQKYEAHVENAVGHEVQRLTVTDL	360
Db	301	TMMDGGSTTTAVAVVEILDANDNAPMDPQKYEAHVENAVGHEVQRLTVTDL	360
QY	361	AMRATYLMGGDGDGHFTTTHPESNQGLITTRKGLDPEAKNQHTLYVEVNEA	420
Db	361	AMRATYLMGGDGDGHFTTTHPESNQGLITTRKGLDPEAKNQHTLYVEVNEA	420
QY	421	PTSTATIVHVEDVNEAPVFPSPKVEVQEGIPTEGPEVCVYTAEDPKENKIS	480
Db	421	PTSTATIVHVEDVNEAPVFPSPKVEVQEGIPTEGPEVCVYTAEDPKENKIS	480
QY	481	DPAGWLAMPDPSGQVTAAGTLDREDEQFVRNNIYEVWLAMDNGSPPTGTG	540
Db	481	DPAGWLAMPDPSGQVTAAGTLDREDEQFVRNNIYEVWLAMDNGSPPTGTG	540
QY	541	VNDHGPVPEPRQITICNQSPVRHLNITDKLSPHTSPPQAQITDSDIYWTAE	600
Db	541	VNDHGPVPEPRQITICNQSPVRHLNITDKLSPHTSPPQAQITDSDIYWTAE	600
QY	601	TVVLSLKFLKQDYDVHLSLDHGKKEQITVIRATVCDCHGHVETCPGPKGG	660
Db	601	TVVLSLKFLKQDYDVHLSLDHGKKEQITVIRATVCDCHGHVETCPGPKGG	660
QY	661	GAVLALLFLLVLLVLLVLRKKRKEPFLLPEDTRDNVFIYEGEGGEDQYDIT	720
Db	661	GAVLALLFLLVLLVLLVLRKKRKEPFLLPEDTRDNVFIYEGEGGEDQYDIT	720
QY	721	GLEAREVVLNVDVATITPTMYRPRANPDEIGNFIENLKAANTDPTAPPD	780
Db	721	GLEAREVVLNVDVATITPTMYRPRANPDEIGNFIENLKAANTDPTAPPD	780
QY	781	DYEGSGDAASLSLSLSSASDQDDYDYNLWNGSRFKKLADMYGGGEDD	829
Db	781	DYEGSGDAASLSLSLSSASDQDDYDYNLWNGSRFKKLADMYGGGEDD	829

RESULT 10
US-10-295-027-896
;; Sequence 896, Application US/10295027
;; Publication No. US20030232350A1
;; GENERAL INFORMATION:
;; APPLICANT: Afar, Daniel
;; APPLICANT: Aziz, Natasha
;; APPLICANT: Ginsberg, Wendy M.
;; APPLICANT: Gish, Kurt C.
;; APPLICANT: Glynn, Richard
;; APPLICANT: Hevezi, Peter A.
;; APPLICANT: Mack, David H.
;; APPLICANT: Murray, Richard
;; APPLICANT: Watson, Susan R.
;; APPLICANT: Eos Biotechnology, Inc.
;; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
;; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
;; FILE REFERENCE: 018501-012500US
;; CURRENT APPLICATION NUMBER: US/10/295,027
;; CURRENT FILING DATE: 2002-11-13
;; PRIOR APPLICATION NUMBER: US 09/663,733
;; PRIOR FILING DATE: 2000-09-15
;; PRIOR APPLICATION NUMBER: US 60/350,666
;; PRIOR FILING DATE: 2001-11-13
;; PRIOR APPLICATION NUMBER: US 60/335,394
;; PRIOR FILING DATE: 2001-11-15
;; PRIOR APPLICATION NUMBER: US 60/332,464
;; PRIOR FILING DATE: 2001-11-21
;; PRIOR APPLICATION NUMBER: US 60/334,393
;; PRIOR FILING DATE: 2001-11-29
;; PRIOR APPLICATION NUMBER: US 60/340,376
;; PRIOR FILING DATE: 2001-12-14
;; PRIOR APPLICATION NUMBER: US 60/347,211
;; PRIOR FILING DATE: 2002-01-08
;; PRIOR APPLICATION NUMBER: US 60/347,349
;; PRIOR FILING DATE: 2002-01-10
;; PRIOR APPLICATION NUMBER: US 60/355,250
;; PRIOR FILING DATE: 2002-02-08
;; PRIOR APPLICATION NUMBER: US 60/356,714
;; PRIOR FILING DATE: 2002-02-13
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 1386
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 896
;; LENGTH: 829
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-295-027-896

Query Match 100.0%; Score 829; DB 15; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 829; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MGLPRGLASLLLLQVCLWQCAASEPCRAVFRFAEAVTLFAGAGAEQPGALGKVFMC	60
Db	1	MGLPRGLASLLLLQVCLWQCAASEPCRAVFRFAEAVTLFAGAGAEQPGALGKVFMC	60
QY	61	QEPALFSTDNDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAPISVP	120
Db	61	QEPALFSTDNDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRDWVAPISVP	120
QY	121	KGFPFQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLILNKP	180
Db	121	KGFPFQRLNQLKSNKDRDTKIFYSITGPGADSPPEGVFAVEKETGWLILNKP	180
QY	181	YELFGHVSNGASVEDPMNIIIVTDQNDHKPKFTQDTRFGSVLEGLVPGTSMQV	240
Db	181	YELFGHVSNGASVEDPMNIIIVTDQNDHKPKFTQDTRFGSVLEGLVPGTSMQV	240
QY	241	DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTHRSTGTSIVSSGLDREKVP	300
Db	241	DEDDAIYTYNGVAVSIHSGEPKDPHDLMTTHRSTGTSIVSSGLDREKVP	300

QY 301 TMDGDSGTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHEVORLTVTDLDPNSP 360
Db 301 TMDGDSGTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHEVORLTVTDLDPNSP 360
QY 361 AWRATYILMGDDGDGHTTITTHPESNOGILTTTRKGLDFAKNQHTLYVEVNEAPVVLK 420
Db 361 AWRATYILMGDDGDGHTTITTHPESNOGILTTTRKGLDFAKNQHTLYVEVNEAPVVLK 420
QY 421 PTSTATIVHVEDVNEAPVFPVPSKVVEQEGIPTGEPVCVYTAEDPDKENQISYRILR 480
Db 421 PTSTATIVHVEDVNEAPVFPVPSKVVEQEGIPTGEPVCVYTAEDPDKENQISYRILR 480
QY 481 DPAGWLAMPDSGOVTAVGLTDREDSQFVRNNIYEVNVLAMDNQSPPTTGTTGLLLTLID 540
Db 481 DPAGWLAMPDSGOVTAVGLTDREDSQFVRNNIYEVNVLAMDNQSPPTTGTTGLLLTLID 540
QY 541 VNRHGPVPEPRQITICNOSPVRHVLNITDKLSPHTSPFQAQLTDDSDIYWTAEVNEEGD 600
Db 541 VNRHGPVPEPRQITICNOSPVRHVLNITDKLSPHTSPFQAQLTDDSDIYWTAEVNEEGD 600
QY 601 TVVLSLKKFLKQDQYDVHLSLSDHGKKEQLTVIRATVCDCHGVETCPGPKGGFLLPVL 660
Db 601 TVVLSLKKFLKQDQYDVHLSLSDHGKKEQLTVIRATVCDCHGVETCPGPKGGFLLPVL 660
QY 661 GAVLALLFLLVLLVLLVLRKKRKEPLLLPEDDTRDNVYFVYEGEGEEDQDYDITQLHR 720
Db 661 GAVLALLFLLVLLVLLVLRKKRKEPLLLPEDDTRDNVYFVYEGEGEEDQDYDITQLHR 720
QY 721 GLEARPEVLRLNDVAPTIITPMYRPRPANDPDIENKAANTDTPAPPYDILLVF 780
Db 721 GLEARPEVLRLNDVAPTIITPMYRPRPANDPDIENKAANTDTPAPPYDILLVF 780
QY 781 DYEGSGDAASLSLTSASDQDDYDYLNEWGRPKKLADMYGGEDD 829
Db 781 DYEGSGDAASLSLTSASDQDDYDYLNEWGRPKKLADMYGGEDD 829

RESULT 11
US-10-229-345-16
; Sequence 16, Application US/10229345
; Publication No. US20040038220A1
; GENERAL INFORMATION:
; APPLICANT: MARKOWITZ, Sanford D.
; TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS
; FILE REFERENCE: CWRU-P01-003
; CURRENT APPLICATION NUMBER: US/10/229,345
; CURRENT FILING DATE: 2002-08-26
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 16
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-229-345-16

Query Match 87.8%; Score 728; DB 12; Length 829;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGLPRGLASLLLLQVCWLQCAASEPCRAVFREAEVTLLEAGGAEQPGALGKVFMGCPG 60
Db 1 MGLPRGLASLLLLQVCWLQCAASEPCRAVFREAEVTLLEAGGAEQPGALGKVFMGCPG 60
QY 61 QEPALFSTNDNDFTVRNGETVQERRSLKERNPLKIPPSKRIILRRHKRDWVAPISVPENG 120
Db 61 QEPALFSTNDNDFTVRNGETVQERRSLKERNPLKIPPSKRIILRRHKRDWVAPISVPENG 120
QY 121 KGFPPRLNQLKSNKDRDRTKIYFSTITGADSPPGVFAVEKETGWLILNKLPLDREEIAK 180
Db 121 KGFPPRLNQLKSNKDRDRTKIYFSTITGADSPPGVFAVEKETGWLILNKLPLDREEIAK 180
QY 181 YELFGHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTRFGSVLEGLVPGTSVMQVTAT 240

Db 181 YELFGHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTRFGSVLEGLVPGTSVMQVTAT 240
QY 241 DEDDAIYTYNGVAVYSHSQEPKDPHLMFTIHRSTGTISVISGLDREKVPYTLTIOA 300
Db 241 DEDDAIYTYNGVAVYSHSQEPKDPHLMFTIHRSTGTISVISGLDREKVPYTLTIOA 300
QY 301 TMDGDSGTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHEVORLTVTDLDPNSP 360
Db 301 TMDGDSGTTTAVAVVEILDANDNAPMFDPOKYEAHVPENAVGHEVORLTVTDLDPNSP 360
QY 361 AWRATYILMGDDGDGHTTITTHPESNOGILTTTRKGLDFAKNQHTLYVEVNEAPVVLK 420
Db 361 AWRATYILMGDDGDGHTTITTHPESNOGILTTTRKGLDFAKNQHTLYVEVNEAPVVLK 420
QY 421 PTSTATIVHVEDVNEAPVFPVPSKVVEQEGIPTGEPVCVYTAEDPDKENQISYRILR 480
Db 421 PTSTATIVHVEDVNEAPVFPVPSKVVEQEGIPTGEPVCVYTAEDPDKENQISYRILR 480
QY 481 DPAGWLAMPDSGOVTAVGLTDREDSQFVRNNIYEVNVLAMDNQSPPTTGTTGLLLTLID 540
Db 481 DPAGWLAMPDSGOVTAVGLTDREDSQFVRNNIYEVNVLAMDNQSPPTTGTTGLLLTLID 540
QY 541 VNRHGPVPEPRQITICNOSPVRHVLNITDKLSPHTSPFQAQLTDDSDIYWTAEVNEEGD 600
Db 541 VNRHGPVPEPRQITICNOSPVRHVLNITDKLSPHTSPFQAQLTDDSDIYWTAEVNEEGD 600
QY 601 TVVLSLKKFLKQDQYDVHLSLSDHGKKEQLTVIRATVCDCHGVETCPGPKGGFLLPVL 660
Db 601 TVVLSLKKFLKQDQYDVHLSLSDHGKKEQLTVIRATVCDCHGVETCPGPKGGFLLPVL 660
QY 661 GAVLALLFLLVLLVLLVLRKKRKEPLLLPEDDTRDNVYFVYEGEGEEDQDYDITQLHR 720
Db 661 GAVLALLFLLVLLVLLVLRKKRKEPLLLPEDDTRDNVYFVYEGEGEEDQDYDITQLHR 720
QY 721 GLEARPEVLRLNDVAPTIITPMYRPRPANDPDIENKAANTDTPAPPYDILLVF 780
Db 721 GLEARPEVLRLNDVAPTIITPMYRPRPANDPDIENKAANTDTPAPPYDILLVF 780
QY 781 DYEGSGDAASLSLTSASDQDDYDYLNEWGRPKKLADMYGGEDD 829
Db 781 DYEGSGDAASLSLTSASDQDDYDYLNEWGRPKKLADMYGGEDD 829

RESULT 12
US-10-274-177-16
; Sequence 16, Application US/10274177
; Publication No. US20040038225A1
; GENERAL INFORMATION:
; APPLICANT: MARKOWITZ, Sanford D.
; TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS
; FILE REFERENCE: CWRU-P01-003
; CURRENT APPLICATION NUMBER: US/10/274,177
; CURRENT FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: US/10/229,345
; PRIOR FILING DATE: 2002-08-26
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 16
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-274-177-16

Query Match 87.8%; Score 728; DB 12; Length 829;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGLPRGLASLLLLQVCWLQCAASEPCRAVFREAEVTLLEAGGAEQPGALGKVFMGCPG 60
Db 1 MGLPRGLASLLLLQVCWLQCAASEPCRAVFREAEVTLLEAGGAEQPGALGKVFMGCPG 60
QY 61 QEPALFSTNDNDFTVRNGETVQERRSLKERNPLKIPPSKRIILRRHKRDWVAPISVPENG 120

```
Db      61  QEPALFSTDNDDFTVRNGETVQERSLKERNPLKIFPSKILRRKRDWVAPISVPENG 120
Qy      121  KGPPFORLNOLKSNKDRDTHKISYITGPGADSPPEGVFAVEKETGWLILNKPLDREETAK 180
Db      121  KGPPFORLNOLKSNKDRDTHKISYITGPGADSPPEGVFAVEKETGWLILNKPLDREETAK 180
Qy      181  YELFGHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240
Db      181  YELFGHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240
Qy      241  DDDALITYNGVAVSISHSQEPKDPHDLMTTHRSTGTISVSSGLDREKPEYTLTQA 300
Db      241  DDDALITYNGVAVSISHSQEPKDPHDLMTTHRSTGTISVSSGLDREKPEYTLTQA 300
Qy      301  TMDGDSGTTTAVAVVEILDANDNAPMFPQKYEAHVPENAVGEVQRLTVDLDAPNSP 360
Db      301  TMDGDSGTTTAVAVVEILDANDNAPMFPQKYEAHVPENAVGEVQRLTVDLDAPNSP 360
Qy      361  AWRATYLMGDDGDHFTTTHPESNQGLTTRKGLDFAKNQHTLYVEVTNEAPFVLKL 420
Db      361  AWRATYLMGDDGDHFTTTHPESNQGLTTRKGLDFAKNQHTLYVEVTNEAPFVLKL 420
Qy      421  PTSTATIVVHVEDVNEAPVFPSPKVEVQEGIPTGEPVCVVTAEPPDKENOKISYRILR 480
Db      421  PTSTATIVVHVEDVNEAPVFPSPKVEVQEGIPTGEPVCVVTAEPPDKENOKISYRILR 480
Qy      481  DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGLLLTLD 540
Db      481  DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGLLLTLD 540
Qy      541  VNDHGVPPEPQITICNQSPVRHLNITDKDLSPTSPFOQLTDDSDIYWTAEVNEEGD 600
Db      541  VNDHGVPPEPQITICNQSPVRHLNITDKDLSPTSPFOQLTDDSDIYWTAEVNEEGD 600
Qy      601  TVVLSLKKFLKQDHYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGGFIPVL 660
Db      601  TVVLSLKKFLKQDHYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGGFIPVL 660
Qy      661  GAVLALLFLLVLLVLLVLRKKRKEPILLPEDDTRDNVYFVYEGEGGEDDYDITQLHR 720
Db      661  GAVLALLFLLVLLVLLVLRKKRKEPILLPEDDTRDNVYFVYEGEGGEDDYDITQLHR 720
Qy      721  GLEARPEVLRNDVAPTIIPTMYRPRPNDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
Db      721  GLEARPEVLRNDVAPTIIPTMYRPRPNDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
Qy      781  DYEAGSDAASLSLSSASDQDDYDYLNWGSRFKKLADMYGGGEDD 829
Db      781  DYEAGSDAASLSLSSASDQDDYDYLNWGSRFKKLADMYGGGEDD 829
```

RESULT 13

```
US-10-301-822-18
; Sequence 18, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MEMO1-029P2RNM
; CURRENT APPLICATION NUMBER: US/10/301,822
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
```

```
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Homo Sapiens
; US-10-301-822-18
```

```
Query March 87.8%; Score 728; DB 14; Length 829;
Best Local Similarity 99.9%; Pred.No. 0;
Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      1  MGLPRGPLASILLIQVCWLQCAASEPCRAVFREAEVTLAAGAEQEPGALGVFMGCPG 60
Db      1  MGLPRGPLASILLIQVCWLQCAASEPCRAVFREAEVTLAAGAEQEPGALGVFMGCPG 60
Qy      61  QEPALFSTDNDDFTVRNGETVQERSLKERNPLKIFPSKILRRKRDWVAPISVPENG 120
Db      61  QEPALFSTDNDDFTVRNGETVQERSLKERNPLKIFPSKILRRKRDWVAPISVPENG 120
Qy      121  KGPPFORLNOLKSNKDRDTHKISYITGPGADSPPEGVFAVEKETGWLILNKPLDREETAK 180
Db      121  KGPPFORLNOLKSNKDRDTHKISYITGPGADSPPEGVFAVEKETGWLILNKPLDREETAK 180
Qy      181  YELFGHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240
Db      181  YELFGHAVSENGASVEDPMNISIIVTDQNDHKPKFTQDTFRGSVLEGLVPGTSVMQVAT 240
Qy      241  DDDALITYNGVAVSISHSQEPKDPHDLMTTHRSTGTISVSSGLDREKPEYTLTQA 300
Db      241  DDDALITYNGVAVSISHSQEPKDPHDLMTTHRSTGTISVSSGLDREKPEYTLTQA 300
Qy      301  TMDGDSGTTTAVAVVEILDANDNAPMFPQKYEAHVPENAVGEVQRLTVDLDAPNSP 360
Db      301  TMDGDSGTTTAVAVVEILDANDNAPMFPQKYEAHVPENAVGEVQRLTVDLDAPNSP 360
Qy      361  AWRATYLMGDDGDHFTTTHPESNQGLTTRKGLDFAKNQHTLYVEVTNEAPFVLKL 420
Db      361  AWRATYLMGDDGDHFTTTHPESNQGLTTRKGLDFAKNQHTLYVEVTNEAPFVLKL 420
Qy      421  PTSTATIVVHVEDVNEAPVFPSPKVEVQEGIPTGEPVCVVTAEPPDKENOKISYRILR 480
Db      421  PTSTATIVVHVEDVNEAPVFPSPKVEVQEGIPTGEPVCVVTAEPPDKENOKISYRILR 480
Qy      481  DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGLLLTLD 540
Db      481  DPAGWLAMPDPSGQVAVGTLDREDEQFVRNNIYEVWVLAMDNQSPPTTGTGLLLTLD 540
Qy      541  VNDHGVPPEPQITICNQSPVRHLNITDKDLSPTSPFOQLTDDSDIYWTAEVNEEGD 600
Db      541  VNDHGVPPEPQITICNQSPVRHLNITDKDLSPTSPFOQLTDDSDIYWTAEVNEEGD 600
Qy      601  TVVLSLKKFLKQDHYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGGFIPVL 660
Db      601  TVVLSLKKFLKQDHYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPKGGFIPVL 660
Qy      661  GAVLALLFLLVLLVLLVLRKKRKEPILLPEDDTRDNVYFVYEGEGGEDDYDITQLHR 720
Db      661  GAVLALLFLLVLLVLLVLRKKRKEPILLPEDDTRDNVYFVYEGEGGEDDYDITQLHR 720
Qy      721  GLEARPEVLRNDVAPTIIPTMYRPRPNDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
Db      721  GLEARPEVLRNDVAPTIIPTMYRPRPNDPDEIGNFIENLKAANTDPTAPPYDTLLVF 780
Qy      781  DYEAGSDAASLSLSSASDQDDYDYLNWGSRFKKLADMYGGGEDD 829
Db      781  DYEAGSDAASLSLSSASDQDDYDYLNWGSRFKKLADMYGGGEDD 829
```

RESULT 14

US-10-295-027-338

; Sequence 338, Application US/10295027

; Publication No. US20030232350A1

; GENERAL INFORMATION:

; APPLICANT: Afar, Daniel

; APPLICANT: Aziz, Natasha

; APPLICANT: Ginsberg, Wendy M.

; APPLICANT: Gish, Kurt C.

; APPLICANT: Glyne, Richard

; APPLICANT: Hevezi, Peter A.

; APPLICANT: Mack, David H.

; APPLICANT: Murray, Richard

; APPLICANT: Watson, Susan R.

; APPLICANT: Eos Biotechnology, Inc.

; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and

; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer

; FILE REFERENCE: 018501-012500US

; CURRENT APPLICATION NUMBER: US/10/295,027

; CURRENT FILING DATE: 2002-11-13

; PRIOR APPLICATION NUMBER: US 09/663,733

; PRIOR FILING DATE: 2000-09-15

; PRIOR APPLICATION NUMBER: US 60/350,666

; PRIOR FILING DATE: 2001-11-13

; PRIOR APPLICATION NUMBER: US 60/335,394

; PRIOR FILING DATE: 2001-11-15

; PRIOR APPLICATION NUMBER: US 60/332,464

; PRIOR FILING DATE: 2001-11-21

; PRIOR APPLICATION NUMBER: US 60/334,393

; PRIOR FILING DATE: 2001-11-29

; PRIOR APPLICATION NUMBER: US 60/340,376

; PRIOR FILING DATE: 2001-12-14

; PRIOR APPLICATION NUMBER: US 60/347,211

; PRIOR FILING DATE: 2002-01-08

; PRIOR APPLICATION NUMBER: US 60/347,349

; PRIOR FILING DATE: 2002-01-10

; PRIOR APPLICATION NUMBER: US 60/355,250

; PRIOR FILING DATE: 2002-02-08

; PRIOR APPLICATION NUMBER: US 60/356,714

; PRIOR FILING DATE: 2002-02-13

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 1386

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 338

; LENGTH: 829

; TYPE: PRT

; ORGANISM: Homo sapiens

; US-10-295-027-338

Query Match 87.8%; Score 728; DB 15; Length 829;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MGLPRGLASLLLLQVWLQCAASEFCRAVFREAEVTLBAGGAEBPGQALGVFWGCPG	60
Db	1	MGLPRGLASLLLLQVWLQCAASEFCRAVFREAEVTLBAGGAEBPGQALGVFWGCPG	60
QY	61	QEPALFSTNDNDDFTVNGETVQERRSLKERNPLKIPPSKRIIRRHKRDWVAPISVPENG	120
Db	61	QEPALFSTNDNDDFTVNGETVQERRSLKERNPLKIPPSKRIIRRHKRDWVAPISVPENG	120
QY	121	KGFPORLQKKNKORDTKIFYSITGPGADSPPEGVFAVEKETGWLNLKPLDREETAK	180
Db	121	KGFPORLQKKNKORDTKIFYSITGPGADSPPEGVFAVEKETGWLNLKPLDREETAK	180
QY	181	YELFGHVAENGASVEDPNNISIVTDQNDHKPKFTQDTFRGSVLPGTSVNVQVAT	240
Db	181	YELFGHVAENGASVEDPNNISIVTDQNDHKPKFTQDTFRGSVLPGTSVNVQVAT	240
QY	241	DEDDALTYTNGVAVSIHQEPKDPHDLMTFTHRSVTGTSVSSGIDREKPEYTLTQA	300
Db	241	DEDDALTYTNGVAVSIHQEPKDPHDLMTFTHRSVTGTSVSSGIDREKPEYTLTQA	300

QY	301	TDMDGGSTTTAVAVVEILDANDNAPMFPDPOKYEAHVPENAVGHVORLTVTDLDAENSP	360
Db	301	TDMDGGSTTTAVAVVEILDANDNAPMFPDPOKYEAHVPENAVGHVORLTVTDLDAENSP	360
QY	361	AWRATYLIMGGDDGDHFTITTHPSNQIGILTRKGLDFEAKNQHTLVVETNEAPFVLKL	420
Db	361	AWRATYLIMGGDDGDHFTITTHPSNQIGILTRKGLDFEAKNQHTLVVETNEAPFVLKL	420
QY	421	PTSTATIVVHVEDVNEAPFVPPSKVVEQSIGITGSPVCVYTAEDPKENQKISYILR	480
Db	421	PTSTATIVVHVEDVNEAPFVPPSKVVEQSIGITGSPVCVYTAEDPKENQKISYILR	480
QY	481	DPAGWLAMPDPSGOVTAAGTLDREDEQFVRNIIYEVWLAMDNQSPPTTGTGTLTLLID	540
Db	481	DPAGWLAMPDPSGOVTAAGTLDREDEQFVRNIIYEVWLAMDNQSPPTTGTGTLTLLID	540
QY	541	VNDHGPVPEPRQITICNQSPVRHLNITDKLSPTSPFQAQLTDDSDIYMTAEVNEEGD	600
Db	541	VNDHGPVPEPRQITICNQSPVRHLNITDKLSPTSPFQAQLTDDSDIYMTAEVNEEGD	600
QY	601	TVVLSLKKFLKQDQDYDVHLSLSHGKKEQLTVIRATVCDCHGHVETCPGPKGGFILLPVL	660
Db	601	TVVLSLKKFLKQDQDYDVHLSLSHGKKEQLTVIRATVCDCHGHVETCPGPKGGFILLPVL	660
QY	661	GAVLALLFLVLLVLLVLRKKRKEIPLLPDDDDTRDNVFFYVYGGEGGEEQDQDITQHR	720
Db	661	GAVLALLFLVLLVLLVLRKKRKEIPLLPDDDDTRDNVFFYVYGGEGGEEQDQDITQHR	720
QY	721	GLEARPEVLENDVAPTIITPMYRPRPANDDEIGNFIIENLKAANTDPTAPPYDTLLVF	780
Db	721	GLEARPEVLENDVAPTIITPMYRPRPANDDEIGNFIIENLKAANTDPTAPPYDTLLVF	780
QY	781	DYEGSGSDAASLSLTSSASDQDQDYDLNWSGRFKKLADMTGGGDD	829
Db	781	DYEGSGSDAASLSLTSSASDQDQDYDLNWSGRFKKLADMTGGGDD	829

RESULT 15
US-10-295-027-1246
; Sequence 1246, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glyne, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10

Search completed: September 21, 2004, 22:38:15
Job time : 145 secs

PRIOR APPLICATION NUMBER: US 60/355,250
PRIOR FILING DATE: 2002-02-08
PRIOR APPLICATION NUMBER: US 60/356,714
PRIOR FILING DATE: 2002-02-13
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1386
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1246
LENGTH: 829
TYPE: PRT
ORGANISM: Homo sapiens
US-10-295-027-1246

Query Match 87.8%; Score 728; DB 15; Length 829;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 828; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGLPRGLASLLLLQVCWLQCAASPCRAVPFAEAVTLEAGGAEQEPGQALGKVFMGCPG 60
Db |||||
QY 1 MGLPRGLASLLLLQVCWLQCAASPCRAVPFAEAVTLEAGGAEQEPGQALGKVFMGCPG 60
Db |||||

QY 61 QEPALFSTNDQFTVANGTQVRRSLKERNPLKIPPSKRILRRHKRDWVAPIISVPENG 120
Db |||||

QY 61 QEPALFSTNDQFTVANGTQVRRSLKERNPLKIPPSKRILRRHKRDWVAPIISVPENG 120
Db |||||

QY 121 KGFPPORLNQKSKNRDRTKIYISITGCGADSPPEGVFAVEKETGKLLINKPLDREETA 180
Db |||||

QY 121 KGFPPORLNQKSKNRDRTKIYISITGCGADSPPEGVFAVEKETGKLLINKPLDREETA 180
Db |||||

QY 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKTQDTFRGSVLEGLVPGTSVMQVTAT 240
Db |||||

QY 181 YELFGHAVSENGASVEDPMNISIIIVTDQNDHKPKTQDTFRGSVLEGLVPGTSVMQVTAT 240
Db |||||

QY 241 DEDDAIYTYNGVAVSIHQEPKDPHDLFTTHRSTGTISVISGLDREKVPETILTQA 300
Db |||||

QY 241 DEDDAIYTYNGVAVSIHQEPKDPHDLFTTHRSTGTISVISGLDREKVPETILTQA 300
Db |||||

QY 301 TMDGSGSTTTAVAVVEILDANDNAPMDFPKYEAHVAVENAVGHEVQRLTVTDLPNSP 360
Db |||||

QY 301 TMDGSGSTTTAVAVVEILDANDNAPMDFPKYEAHVAVENAVGHEVQRLTVTDLPNSP 360
Db |||||

QY 361 AWRATYLMGGDDGDHFTITTHPESNQILTRKGLDPEAKNQHTLYEVNTNEAPFVLKL 420
Db |||||

QY 361 AWRATYLMGGDDGDHFTITTHPESNQILTRKGLDPEAKNQHTLYEVNTNEAPFVLKL 420
Db |||||

QY 421 PTSTATIVVHVEDVNEAPVFPSPKVEVQEGIPTEGVCVYTAEDPDKENOKISYRILR 480
Db |||||

QY 421 PTSTATIVVHVEDVNEAPVFPSPKVEVQEGIPTEGVCVYTAEDPDKENOKISYRILR 480
Db |||||

QY 481 DPAGWLAMPDPSGQVAVGTDLREDEQFVRNNIYEVWVLAMDNAGSPPTTGTGLTLTLD 540
Db |||||

QY 481 DPAGWLAMPDPSGQVAVGTDLREDEQFVRNNIYEVWVLAMDNAGSPPTTGTGLTLTLD 540
Db |||||

QY 541 VNDHGVPEPRQITICNQSPVRHVNITDKDLSPTSPFOALITDDSDIYWTAEVNEEGD 600
Db |||||

QY 541 VNDHGVPEPRQITICNQSPVRHVNITDKDLSPTSPFOALITDDSDIYWTAEVNEEGD 600
Db |||||

QY 601 TVVLSLKKFLKQDYDVHLSLSDHGNKEQTVIRATVCDCHGHVETCPGPKGGFLLPVL 660
Db |||||

QY 601 TVVLSLKKFLKQDYDVHLSLSDHGNKEQTVIRATVCDCHGHVETCPGPKGGFLLPVL 660
Db |||||

QY 661 GAVLALLFLLVLLVLLVRKKRIKEPLLPEDDTRDNVFFYEGGEGEDQDYDTQLHR 720
Db |||||

QY 661 GAVLALLFLLVLLVLLVRKKRIKEPLLPEDDTRDNVFFYEGGEGEDQDYDTQLHR 720
Db |||||

QY 721 GLEAPVWLRNDVAPTIIPTRMYRPRPANPDEIGNFIENLKAANTDTAPPYDTLTVF 780
Db |||||

QY 721 GLEAPVWLRNDVAPTIIPTRMYRPRPANPDEIGNFIENLKAANTDTAPPYDTLTVF 780
Db |||||

QY 781 DYEGSGSDAASLSLSSASDQDQDYDYLNEWGSRFKKLADMYGGGEDD 829
Db |||||

QY 781 DYEGSGSDAASLSLSSASDQDQDYDYLNEWGSRFKKLADMYGGGEDD 829
Db |||||

This Page Blank (uspto)